Vol. 53 DECEMBER 1962 No. 6

Journal of Educational Psychology

CONTENTS

Organizer, General Background, and Antecedent Learning Variables in Sequential Verbal Learning 243
DAVID P. AUSUBEL AND DONALD FITZGERALD

The Effect of Intermittent Confirmation in Programed Instruction 250
JOHN D. KRUMBOLTZ AND RONALD G. WEISMAN

Cross-Cultural Studies of Teenagers' Problems 254
H. H. REMMERS

Parental Identification in Relation to Vocational Interest Development 262
JOHN O. CRITES

Responses of Bright, Normal, and EMH Children to an Orally-Administered Children's Manifest Anxiety Scale 271
NEIL A. CARRIER, KENNETH D. ORTON, AND LESLIE F. MALPASS

Changes in Student Teachers' Attitudes toward Childhood Behavior Problems 275
GEORGE A. PINCKNEY

Discernment of Teacher Characteristics by TAT Sequence Analysis 279
M. INNOCENTIA BURKARD

The Reliability of Measures of Handwriting Quality 288
LEONARD S. FELDT

Failure Imagery in the Fantasy of Eighth Graders as a Function of Three Conditions of Induced Arousal 293
RICHARD C. ANDERSON

Test Anxiety, Experimental Instructions, and the Wechsler Adult Intelligence Scale 299
IRWIN G. SARASON AND JAMES MINARD

Influences on the Student's Motivation to Seek Advanced Training: Another Look 303
ALEXANDER W. ASTIN

Rival Hypotheses for Explaining the Effects of Different Learning Environments 310
DONALD L. THISTLETHWAITE

Sex, Social Class, and Anxiety as Sources of Variation in School Achievement 316
BEEMAN N. PHILLIPS

This is the last issue of Volume 53.
Volume Title Page and Contents of the Volume appear herein.

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF EDUCATIONAL PSYCHOLOGY

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

JOHN B. CARROLL
Harvard University
LEE J. CRONBACH
University of Illinois
FRANCIS J. DIVESTA
Syracuse University
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
DOROTHEA MCCARTHY
Fordham University
CARSON MCGUIRE
University of Texas
PAUL H. MUSSEN
University of California, Berkeley
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
GEORGE G. THOMPSON
Ohio State University
BENTON J. UNDERWOOD
Northwestern University

The *Journal of Educational Psychology* publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*.

Manuscripts must be accompanied by an abstract of 100–120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR, Managing Editor
ELIZABETH S. REED, Advertising Manager
JOYCE J. CARTER, Editorial Assistant

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the tenth of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $10.00 (Foreign $10.50). Single copies, $2.00.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W., Washington 6, D. C.
Second-Class postage paid at Washington, D. C. and additional offices

Copyright, 1962 by the American Psychological Association, Inc.

JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 53 December 1962 Number 6

ORGANIZER, GENERAL BACKGROUND, AND ANTECEDENT LEARNING VARIABLES IN SEQUENTIAL VERBAL LEARNING

DAVID P. AUSUBEL AND DONALD FITZGERALD
Bureau of Educational Research
University of Illinois

An analysis of covariance design was used with undergraduate students in studying the effects of an advance organizer, antecedent learning, and general background knowledge on the learning and retention of 2 unfamiliar sequential passages about endocrinology. The organizer, by providing relevant ideational anchorage, suggestively enhanced the learning of the material for those Ss with relatively poor verbal ability. Knowledge of the 1st passage constituted a statistically significant limiting condition in learning the 2nd passage when the influence of both verbal ability and general endocrinological knowledge was statistically controlled. Finally, general background knowledge in endocrinology significantly facilitated the learning and retention of the 1st passage, presumably by increasing the familiarity of the new material.

Typically in the course of meaningful school learning, the learner is introduced to new materials which are variants of and hence relatable to previously learned concepts already established in his cognitive structure. Ausubel and Fitzgerald (1961) have shown that under these circumstances the learner's ability to discriminate between the two sets of concepts has important implications for learning and retention. They demonstrated that such discriminability is partly a function of the clarity and stability of the previously learned concepts, and that when discriminability is low because of inadequate prior knowledge, learning and retention can be enhanced by the use of "comparative organizers."

A somewhat different learning situation, however, not uncommonly arises when, for various reasons, students are required to learn a new unit of sequentially organized material which cannot be related to previously learned concepts. No specific referents, either explicit or implicit, are therefore available for this material in cognitive structure. Organizers, under these conditions, obviously cannot be used to increase discriminability; they can at best furnish ideational anchorage.

This type of learning situation also provides an excellent model and test of the transfer of training paradigm in sequential meaningful learning. When subject matter is organized sequentially (so that Part II is built upon and presupposes knowledge of Part I) the student's grasp of the earlier appearing material becomes a crucial factor influencing his learning of the later appearing material. When materials are sequentially presented, therefore, the residue of the earlier appearing material in cognitive structure—i.e., its clarity, stability, and organization—constitutes the impact

of the "antecedent condition" in the transfer of training paradigm. It becomes in effect the anchoring post or ideational scaffolding in cognitive structure for the learning of the later appearing material. If this ideational scaffolding is clear, stable, and well organized, it is reasonable to suppose that it provides better anchorage for new learning and retention than if it is unclear, unstable, and poorly organized.

In addition to furnishing a theoretical model for sequential meaningful learning in terms of the familiar transfer of training paradigm, an analysis of covariance design utilizing this type of learning situation makes possible a more explicit test of the proposition that the quality of Part I learning actually influences Part II learning outcomes when subject matter is sequentially organized. Typically this proposition is assumed to be axiomatically true in the absence of proof because of the difficulty of controlling such significant variables as relative learning ability (which accounts for much of the positive relationship between Part I and Part II learning outcomes), and the availability in cognitive structure of relevant concepts relatable to and influencing the learning of Part II materials which were introduced prior to the Part I materials. In this study the influence of learning ability is held constant by statistical means, and the contaminating influence of antecedent conditions prior to those established by the Part I materials is eliminated both by the use of wholly unfamiliar materials and by statistical control of general background knowledge.

This type of research design also makes it possible to test the effect of general background knowledge in the same subject matter area on the learning of explicitly unfamiliar new material. It is conceivable that the general level of sophistication can affect the learning of new sequential material—not by influencing its discriminability from or relatability to cognitive structure, but more indirectly by increasing its familiarity and hence the learner's feeling of self-confidence in handling it.

It is hypothesized, first, that the expository organizer differentially enhances the learning and retention of completely unfamiliar material for those subjects in the experimental group who have less verbal ability, and who therefore tend spontaneously to structure new verbal material less efficiently and successfully. Second, it is hypothesized that as a result of the enhanced learning of Part I materials achieved by some members of the experimental group with the aid of the organizer in the sequential learning situation, the experimental group as a whole manifests superior learning and retention of Part II materials. Third, it is hypothesized that there is a significantly positive relationship between Part I and Part II retention scores in both experimental and control treatment groups, and that this relationship remains significantly positive even when the influence of learning ability and general background knowledge is eliminated. Finally, it is hypothesized that general background knowledge increases the learning and retention of Part I materials.

METHOD

Subjects

The experimental population consisted of predominantly senior undergraduate students (91 women and 52 men) in six sections of an educational psychology course at the University of Illinois. All subjects were enrolled in 1 of 10 teacher education curricula at the secondary school level. The experiment was conducted separately in each section as a required laboratory exercise, and was performed during regularly scheduled class hours. In order to maximize ego involvement, subjects were informed that after the data were processed their individual scores, as well as the class results, would be reported to them.

Learning Passages, Organizers, and Measuring Instruments

The learning material used in this study consisted of two specially prepared, sequentially related passages dealing with the endocrinology of pubescence. The first passage (approximately 1,400 words) was concerned with the specific hormonal factors initiating and regulating pubescence. The second passage (approximately 1,600 words) considered pathological variations in pubescence and their treatment.

This particular topic was selected because it was anticipated that it would be almost completely unfamiliar to undergraduate students in teacher education. Inasmuch as it dealt with new and rather esoteric ideas at the frontiers of medical physiology, we were confident that nothing in the previous academic background of our subjects had left a residue of established concepts in cognitive structure to which this material was specifically relatable. It was important to satisfy this criterion for two reasons: First, since we were endeavoring to make the meaningfulness and comprehensibility of the second passage as dependent as possible upon knowledge of the first passage, it was necessary to make sure that the later appearing passage in the sequence was not relatable to existing concepts of *other* origin within cognitive structure (i.e., of origin other than the earlier appearing passage). Second, since the organizer in this experiment, unlike the organizer in another recent study (Ausubel & Fitzgerald 1961), was designed merely to provide ideational anchorage for the first passage, and not to increase its discriminability from analogous concepts within the subjects' cognitive structure, it was necessary to avoid any possibility of relatedness between the learning passage and established concepts. The use of unfamiliar learning material also made it possible for all subjects to start from approximately the same base line in learning the two passages.

Empirical confirmation of the unfamiliarity of the pubescence material was obtained when a group of comparable naive subjects, who had not studied the learning passages, made scores which, on the average, were only slightly and not significantly better than chance.

Knowledge of the material in each pubesence passage was tested by a 36-item multiple-choice test. The corrected split-half reliabilities of the two tests were .83 and .74, respectively, and the correlation between them was .51 for the total group of subjects. Scores on both tests showed a satisfactory range of variability, and their distribution did not deviate significantly from the normal distribution curve. Since they were intended as power tests, no time limit was imposed.

The experimental group studied on two separate occasions a 500-word introductory passage (expository *organizer*) which provided an organizational framework for the first passage in terms of the different kinds of uniformity and variability prevailing among the primary and secondary sex characteristics. It was pitched at a much higher level of abstraction, generality, and inclusiveness than the pubescence material itself. The control group also studied on two separate occasions a 500-word introductory passage dealing with uniformity and variability among different cultures in *behavioral* aspects of pubescence. This was intended solely as a control treatment and had no organizational properties whatsoever in relation to the first learning passage. No information was included in either of the introductory passages that could constitute a direct advantage in answering questions on the pubescence material. It was methodologically important to provide a control (nonorganizing) introduction for the control group in order that any obtained differences in retention outcomes between the experimental and control groups be attributed to the organizing properties of the experimental introduction rather than to its presence per se (i.e., the Hawthorne effect).

To test the effect of general background knowledge on the learning of the pubescence material a 20-item multiple-choice General Endocrinology Test (GET) was used. The distribution of scores on this test did not deviate significantly from the normal distribution curve, and the corrected split-half reliability was .88. As a power test, this test also had no time limit.

Scores on the verbal portion of the School and College Ability Test (SCAT), reflective of verbal ability, were available for all subjects.

Procedure

On the first day of experimentation, all subjects first took the GET and then (after assignment to a treatment group) studied either the organizer or control introduction for 8 minutes. Membership in a given treatment group was determined by random assignment. The population of each of the three treatment groups was also stratified by sex so that the proportion of men to women subjects would be the same in each group. This was necessary because of the women's significantly higher verbal SCAT scores. It was possible to administer the two treatments simultaneously because they consisted of identical appearing introductory pas-

sages (with identical sets of directions) differing only in content.

The principle of random assignment to treatment groups was relied upon to equate experimental and control groups with respect to relative learning ability and background knowledge in the pubescence material. To further equate the treatment groups with respect to the latter variable, the relatively few biology majors in the group were identified and divided equally among the experimental and control groups. In any case, however, the influence of background knowledge was minimized by the esoteric nature of the material and its almost complete unfamiliarity to naive subjects (see above).

In order to control for the effects of classroom climate, instructor, and situational variables in the six sections, students *within each* section were equally divided among the two treatment groups. Since homogeneity of variance prevailed, both between sections and treatment groups, for both the pubescence test scores and the GET scores, it was considered justifiable to pool the scores from the various sections and to treat the scores of the two treatment groups on each of these instruments as comparable random samples from the same large population.

Two days after the first experimental session, both treatment groups restudied their respective introductory passages for 6 minutes and then studied the first pubescence passage for 25 minutes. The test on the latter passage came 48 hours later. The second pubescence learning passage was administered 3 days after the test on the first passage. Subjects were given 27 minutes to read and study the second passage and were tested on it 4 days later.

RESULTS AND DISCUSSION

Effect of Organizer on Initial Learning

Comparison of the means of the total experimental and control groups (Table 1) shows that the organizer facilitated the learning and retention of the first pubescence passage ($p = .07$). Analysis of covariance similarly indicates that the overall effect of the organizer on retention scores was not quite significant at the .05 level (Table 2). As hypothesized, however, practically all of the obtained near significant difference between experimental and control groups was derived from the subjects in the lower third of the distribution of verbal ability scores. Experimental subjects in the lower third subgroup of SCAT scores learned and retained significantly more material than did control subjects in the lower third subgroup ($p < .01$), whereas differences between experimental and control subjects in the upper third and middle third subgroups were negligible.

The differential facilitating effect of the organizer on the learning and retention of the first pubescence passage for the subjects with relatively low verbal ability is also revealed by the organizer's leveling influence on the significantly positive relationship between SCAT and retention scores (Tables 2 and 3). The

TABLE 1

ADJUSTED MEAN RETENTION SCORES OF EXPERIMENTAL AND CONTROL GROUPS OF FIRST PUBESCENCE PASSAGE ANALYZED BY VERBAL ABILITY LEVELS

Verbal ability (SCAT)	Treatment group	
	Organizer	Control
Upper third	20.9	21.1
Middle third	19.1	18.5
Lower third	18.3	15.0
Total	19.4	18.2

TABLE 2

COVARIANCE ANALYSIS OF FIRST PUBESCENCE TEST SCORES × SCAT LEVELS, ADJUSTING FOR GET SCORES

Source	Adjusted		
	df	*MS*	*F*
Treatment	1	50	2.27
SCAT level	2	199	9.05*
Interaction	2	37	1.68
Between	5	102	4.64*
Within	125	22	

* .01 level of significance.

TABLE 3

PARTIAL CORRELATIONS BETWEEN SCAT, GET, PUBESCENCE TEST I, AND PUBESCENCE TEST 2

Correlation	Held constant	Treatment group	
		Organizer	Control
GET × Test I	SCAT	.31*	.11
SCAT × Test I	GET	.19*	.50**
GET × Test II	SCAT	.33**	.14
GET × Test II	Test I	.29*	.13
SCAT × Test II	GET	.24	.43**
SCAT × Test II	Test I	.25	.22
Test 1 × Test II	SCAT	.26*	.47**
Test 1 × Test II	GET	.22	.57**

* .05 level of significance.
** .01 level of significance.

partial correlation between SCAT and retention scores, with GET scores held constant, was .50 ($p < .01$) for the control group but only .19 for the experimental group (Table 3). Thus by differentially raising the retention scores of the lower ability subjects, the organizer reduced the intrinsic relationship between SCAT and retention scores.

Suggestive evidence therefore exists for believing that organizers, by providing ideational anchorage, facilitate the learning and retention of totally unfamiliar material for those subjects who have relatively little verbal ability. Subjects of average and better ability are evidently capable of *spontaneously* organizing new learning material around relevant, more inclusive concepts, and hence derive little or no benefit from introduced advance organizers. This kind of interaction between verbal ability and the effect of organizers did not occur in a previous study where the learning material (Buddhism) was more explicitly relatable to previously learned concepts (Christianity) in cognitive structure (Ausubel & Fitzgerald, 1961).

It is possible that the potential facilitating impact of the organizer on the learning and retention of the first pubescence passage was not fully realized in this study because subjects, due to inadequate background knowledge in the immediate subject matter area, experienced difficulty in relating the organizer to existing cognitive structure, and hence in learning it. Obviously, if the organizer itself cannot be firmly anchored to cognitive structure, it cannot serve as a very effective anchoring post. If possible, therefore, in preparing organizers for an unfamiliar field of knowledge, an effort should be made to formulate the organizing principles in terms of concepts that are already familiar to the learner and established in his cognitive structure. The more unfamiliar and esoteric a given learning task is, however, the more difficult it is to implement this precept.

Effect of Organizers on Sequential Learning

It had been hypothesized that the organizer, by virtue of facilitating retention scores on Test I, would thereby enhance the learning and retention of the second pubescence passage. Comparison of the means of the total organizer and control groups (Table 4), as well as reference to the covariance analysis of second pubescence test scores by Test I scores and treatment, holding either

TABLE 4

ADJUSTED MEAN RETENTION SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON SECOND PUBESCENCE PASSAGE ANALYZED BY PUBESCENCE TEST I LEVELS

Pubescence Test I scores	Treatment group	
	Organizer	Control
Upper third	23.2	24.0
Middle third	19.2	21.0
Lower third	19.2	18.1
Total	20.5	21.0

TABLE 5

COVARIANCE ANALYSIS OF SECOND PUBESCENCE TEST SCORES × TEST I LEVELS, ADJUSTING FOR GET SCORES

Source	Adjusted		
	df	*MS*	*F*
Treatment	1	7	.39
Test I levels	2	282	15.7**
Interaction	2	11	.61
Between	5	99	5.50*
Within	131	18	

* .01 level of significance.
** .001 level of significance.

GET (Table 5) or SCAT (Table 6) constant, show that the hypothesis was not confirmed. This finding, of course, was hardly unexpected in view of the relatively slight and only suggestive facilitating effect of the organizer on initial learning. Nevertheless, the significant partial correlations in the experimental group both between GET and Test II scores ($p < .01$), with the influence of SCAT eliminated, and between GET and Test II scores ($p < .05$), with the influence of Test I eliminated, when contrasted with the nonsignificant corresponding correlations in the control group (Table 3), suggest the occurrence of a positive interaction between the effects of the organizer and of general background knowledge. The organizer apparently helped the subjects utilize their existing background knowledge more effectively in structuring the second pubescence passage.

TABLE 6

COVARIANCE ANALYSIS OF SECOND PUBESCENCE TEST SCORES × TEST I LEVELS, ADJUSTING FOR SCAT SCORES

Source	Adjusted		
	df	*MS*	*F*
Treatment	1	12	.67
Test I levels	2	171	9.50*
Interaction	2	6	.34
Between	5	73	4.06*
Within	131	18	

* .01 level of significance.

In order for organizers to be really effective in enhancing sequential learning and retention, it would probably be necessary to use another organizer prior to the second passage. The second organizer would attempt both to provide specific ideational scaffolding for the second passage, and also to increase discriminability between confusable concepts in the two passages.

Effect of Initial Learning on Sequential Learning

As hypothesized, Part I retention scores (initial learning) had a significant facilitating effect on Part II retention scores (sequential learning). Both experimental and control subjects in the upper part of the distribution of scores on the first pubescence test made higher retention scores on the second pubescence test than did subjects in the lower part of the distribution of first pubescence test scores. This facilitating effect of antecedent learning ($p < .001$; $p < .01$) was manifested when either the influence of general background knowledge (Table 5) or of verbal ability (Table 6) was statistically controlled in a covariance design. Stability and clarity of knowledge of the first passage therefore constituted a significant limiting condition with respect to learning and retaining the material in the second passage.

Sequential organization of subject matter, therefore, can be very effective because each new increment of knowledge serves as an anchoring post for subsequent learning. This presupposes, of course, that the preceding step is always clear, stable, and well organized. If it is not, the learning of all subsequent steps is jeopardized. Hence new material in the sequence should never be intro-

duced until all previous steps are thoroughly mastered. Perhaps the chief pedagogic advantage of the teaching machine lies in its ability to control this crucial variable in sequential learning.

Effect of General Background Knowledge on Initial Learning

The data clearly support the hypothesis that general background knoweldge in endocrinology facilitates the learning of unfamiliar material in the same subject matter field (i.e., pubescence). This facilitating effect was significant at the .05 level when verbal ability was statistically controlled in a covariance design (Table 7). Significant interaction ($p < .03$) also occurred between the organizer and general endocrinology knowledge. Reference to Table 8 shows that the interactional effect was positive for the subjects in the upper third and lower third subgroups of the distribution of GET scores and negative for subjects in the middle third subgroup. The overall interactional effect, however, was positive as shown by the contrast between the significant partial correlation ($p < .05$) in the experimental group between GET and intial retention scores, with SCAT held constant, and the nonsignificant corresponding correlation in the control group (Table 3). Similar positive correlations were reported above in relation to Test I scores.

TABLE 7

COVARIANCE ANALYSIS OF SECOND PUBESCENCE TEST SCORES × GET LEVELS ADJUSTING FOR SCAT SCORES

Source	Adjusted df	MS	F
Treatment	1	33	1.57
GET level	2	76	3.64*
Interaction	2	98	4.67*
Between	5	74	3.52**
Within	131	21	

* .05 level of significance.
** .01 level of significance.

TABLE 8

ADJUSTED MEAN RETENTION SCORES OF EXPERIMENTAL AND CONTROL GROUPS ON FIRST PUBESCENCE PASSAGE ANALYZED BY GET LEVELS

General endocrinology knowledge	Treatment group	
	Organizer	Control
Upper third	21.1	18.3
Middle third	18.4	20.8
Lower third	18.6	16.1
Total	19.4	18.4

General endocrinology knowledge was too distantly related to the pubescence material to provide relevant ideational anchorage. For the most part, it probably facilitated the learning and retention of the pubescence passage by increasing its general familiarity, and hence the subjects' self-confidence in handling it. However, the positive interaction between the effects of the organizer and of general background knowledge in endocrinology suggests that the organizer better enables the subjects to put their background knowledge to effective use in structuring the unfamiliar new material. This finding was consistent for both initial and sequential learning.

REFERENCE

AUSUBEL, D. P., & FITZGERALD, D. The role of discriminability in meaningful verbal learning and retention. *J. educ. Psychol.*, 1961, **52**, 266–274.

(Received April 13, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 250–253

THE EFFECT OF INTERMITTENT CONFIRMATION IN PROGRAMED INSTRUCTION [1]

JOHN D. KRUMBOLTZ AND RONALD G. WEISMAN
Stanford University *Michigan State University*

The effect of intermittent confirmation was tested on 121 students by omitting various patterns of confirming answers from a programed textbook on educational measurement. The schedules included 4 levels of fixed-ratio confirmation and 2 of variable-ratio confirmation. Results based on criterion measures consisting of (a) errors made on the program and (b) performance on a posttest indicated: a negative linear relationship between the number of errors made on the program and the percentage of confirmation provided, no significant effects on the posttest from the various proportions of confirmation, and no evidence of differential effect between fixed-ratio and variable-ratio confirmation on either criterion.

Intermittent reinforcement has the well known effect of prolonging extinction responding in simple response sequences (Lewis, 1960). Experimental tests of the effects of intermittent reinforcement in complex, meaningful learning are difficult to devise because of the problem of controlling the "reinforcement." Several recent attempts to devise such experiments have been reviewed by Krumboltz (1961).

In programed instruction knowledge of results may be considered a form of reinforcement. In many programed textbooks the correct answer to a frame is placed on the page following the frame. A student can confirm each of his answers by comparing it with the correct answer on the following page. Omitting certain of these answers provides an approach for investigating the effects of various schedules of confirmation.

This approach cannot provide a completely adequate way to assess the effects of partial reinforcement because there are certainly other less explicit reinforcers in most programs. Occasionally the correct answer is stated or implied in succeeding frames. Or if the preceding explanation has been adequate, a student may report that he knew he was correct even in the absence of the confirming answer (and sometimes even when a typographical error has made the confirming answer incorrect). Nevertheless, the most prominent reinforcer is the knowledge of results provided by the confirming answer. Providing confirming answers on an intermittent schedule during learning of programed material may increase retention in the same manner that intermittent reinforcement prolongs extinction responding in less complex learning tasks.

METHOD

Subjects

The subjects were drawn from 140 students enrolled in four sections of an introductory course in educational psychology during the summer of 1961. Due to absences the final sample size was reduced to 121. Within each section, the subjects were randomly assigned to each of the six reinforcement schedules.

Materials

The 177-frame programed textbook in this experiment was designed to teach prospective teachers how to interpret educational test results. It covered material on percentiles, norms, means and medians, standard deviations, and

[1] This research was supported by a grant from the United States Office of Education, Educational Media Branch, under Title VII of the National Defense Education Act.

z scores. The linear constructed-response program was not intended to develop computational proficiency but only to develop a useful understanding. The program had previously been administered to 80 students and revised in accordance with observed error rates and performance on a criterion test.

The 50-item completion-type criterion test was also revised after an initial administration in accordance with an item analysis and changes in the program itself. Its estimated corrected split-half reliability was found to be .85.

Procedure

The same basic program was used with all six treatment groups. The only differences were in the number and order of confirming answers provided by the programed textbook. The six variations were as follows:

1. Continuous confirmation: answers for every response were provided.
2. Fixed-ratio 67% confirmation: the confirming answer to every third frame was omitted.
3. Variable-ratio 67% confirmation: on the average only two-thirds of the frames at irregular intervals were followed by confirming answers.
4. Fixed-ratio 33% confirmation: the confirming answer to every third frame only was included.
5. Variable-ratio 33% confirmation: on the average one-third of the frames at irregular intervals were followed by confirming answers.
6. No confirmation: all answers were eliminated.

The above schedules were applied beginning at the eleventh frame in the 177-frame program. Each of the first 10 frames under all six schedules was followed by confirming answers so that students would learn the usual procedure for using programed textbooks.

To prevent generalizing from just one specific set of omitted answers, the particular answers omitted were varied in both the fixed-ratio and variable-ratio booklets. One-third of the fixed-ratio booklets were constructed to coincide with each of the three possible patterns of omitted answers at each of the two reinforcement rates. A table of random numbers was used to prepare 10 different versions of the variable-ratio schedule at each confirmation rate. Modifications in the program were accomplished by inking out the designated answers with felt-tip pens.

Booklets representing each of the six schedules were arranged in random order and handed out to members of each class. Students were given sufficient class time (4 hours) to complete all 177 frames in class. All students were instructed to record their answers on a separate sheet of paper provided by the experimenters.

As soon as a student had completed the programed textbook, he was given the 50-item completion-type criterion test. Two criterion measures were analyzed: (*a*) number of errors made on the program itself, and (*b*) number of correct answers on the criterion test.

RESULTS

A summary of the results appears in Table 1. Various tests of hypotheses based on these data were made to answer the following questions:

1. Ignoring the variable-ratio schedules for the moment, do the remaining four schedules (all of which may be

TABLE 1

SUMMARY OF RESULTS FOR EACH SCHEDULE OF CONFIRMATION

Schedule of confirmation	*N*	Number of errors in 177-frame program		Score on criterion test		Correlation between criterion test and grade point average
		M	*SD*	*M*	*SD*	
Continuous	20	19.1	14.2	32.3	6.7	.62
Fixed-ratio 67%	20	19.7	12.6	32.7	5.4	−.10
Variable-ratio 67%	18	19.2	9.0	33.0	7.8	.42
Fixed-ratio 33%	20	23.5	15.3	30.9	6.0	.58
Variable-ratio 33%	18	26.5	12.6	32.6	6.9	.61
No	25	33.7	16.3	31.1	6.6	.57

considered fixed-ratio) cause different error rates on the program itself? Significant differences were obtained ($F=4.61$; $df=3/81$; $p < .005$). Furthermore, a test of the linearity of the trend over the four degrees of confirmation, 0%, 33%, 67%, and 100%, revealed a significant linear trend ($F=11.68$, $df=1/81$; $p < .001$). That is, there is a significant negative relationship between the amount of confirmation and the number of errors made on the program itself; the more confirmation, the fewer program errors.

2. Do these same four schedules produce different scores on the criterion test? A similar analysis of the criterion test scores resulted in a nonsignificant F less than 1. Thus, the four degrees of confirmation in fixed-ratio schedules did not produce criterion test means sufficiently different to reject the null hypothesis. An attempt was made to gain further precision by computing an analysis of covariance by adjusting the mean criterion test scores for initial differences in students' all-university grade point averages. The covariance analysis was abandoned, however, when the assumption of homogeneous regression was not found tenable ($F=3.48$; $df=3/81$; $p < .05$). As seen in Table 1, all correlations were approximately the same except for the one slightly negative relationship in the Fixed-ratio 67% group. No logical explanation for this discrepancy is apparent.

3. What main and interaction effects result from differences between type of ratio and percentage of confirmation? Here the Continuous confirmation and No confirmation groups were ignored, and a 2×2 analysis of variance of the other four groups was calculated on the program errors (Table 2) and the criterion test scores (Table 3). Differences between fixed-ratio and variable-ratio schedules produced nonsignificant differences on both criteria as well as on the interactions with the percentage of confirmation. No evidence was obtained with the criteria used here on the differential effectiveness of fixed-ratio and variable-ratio schedules of confirmation in programed instruction.

TABLE 2

ANALYSIS OF VARIANCE OF ERRORS ON PROGRAM

Source	*df*	*MS*	*F*
Percentage of confirmation	1	574.75	3.39
Type of ratio	1	15.15	< 1
Interaction	1	22.08	< 1
Within	72	169.27	

TABLE 3

ANALYSIS OF VARIANCE OF CRITERION TEST SCORES

Source	*df*	*MS*	*F*
Percentage of confirmation	1	26.30	< 1
Type of ratio	1	15.09	< 1
Interaction	1	49.18	1.04
Within	72	47.29	

DISCUSSION

The results of the present study did include one effect common to intermittent reinforcement: an increase in error rate during learning under the higher ratio scredules. Groups receiving fewer confirmations made more errors during learning although the difference in error rate was relatively small indicating that considerable confirmation was probably present within the frames themselves.

Thus, while it is clear that confirmation had some effect on the error rate, it is also clear that sequencing and cueing within the program itself were important. It is hypothesized that the effects of various schedules should become increasingly evident as more of the inter-

nal cueing is removed from the program. If confirmed, such a hypothesis would be consistent with some evidence supporting the hypothesis of Evans, Glaser, and Homme (1960) that the relevance of variables such as response mode and immediacy of confirmation is inversely related to the probability of correct responding. For example, Goldbeck (1960) discovered an interaction between response mode and the difficulty level of a "discrete discourse" program. See Krumboltz (1961) for a discussion of the related evidence.

The results of the criterion test did not give evidence that intermittent confirmation improved retention. It is possible, however, that an immediate retention test may not be sensitive to differences in schedules. Krumboltz and Weisman (1962) found that a retention test delayed 2 weeks was more sensitive to the effect of overt responding. This would argue that future experiments should include delayed measures of retention as well as immediate criteria.

In any future experiment along this line additional precision could be obtained by arranging the schedules of confirmation separately for each concept. Many concepts and their interrelationships were taught in the present program, and the schedule of confirmations ignored the particular concept emphasized in each frame. Programs could be arranged so that various schedules would be tested for each major concept regardless of that concept's order in total program.

REFERENCES

EVANS, J., GLASER, R., & HOMME, L. The development and use of a "standard program" for investigating programed verbal learning. *Amer. Psychologist*, 1960, **15**, 424. (Abstract)

GOLDBECK, R. A. The effect of response mode and learning material difficulty on automated instruction. Technical Report No. 1, 1960, American Institute for Research, Santa Barbara, California, Contract No. Nonr-3077(00).

KRUMBOLTZ, J. D. Meaningful learning and retention: Practice and reinforcement variables. *Rev. educ. Res.*, 1961, **31**, 535–546.

KRUMBOLTZ, J. D., & WEISMAN, R. G. The effect of overt versus covert responding to programed instruction on immediate and delayed retention. *J. educ. Psychol.*, 1962, **53**, 89–92.

LEWIS, D. J. Partial reinforcement: A selective review of the literature since 1950. *Psychol. Bull.*, 1960, **57**, 1–28.

(Received December 8, 1961)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 254–261

CROSS-CULTURAL STUDIES OF TEENAGERS' PROBLEMS [1]

H. H. REMMERS
Purdue University

This paper reports findings from representative stratified samples of more than 5000 teenagers in school in the United States, Puerto Rico, West Germany, and India who responded to the Science Research Associates (SRA) Youth Inventory, a problems check list adapted to each of the cultures surveyed. Mean scores, reliability estimates, intercorrelations of subscale scores, and factor analyses of these matrices all lead to the conclusion that (a) the measuring instrument is highly reliable, (b) teenagers' self-perceived problems can be comparably measured across widely different cultures, (c) rankings of problem areas across cultures are highly similar, (d) health problems are of least concern and post-high-school problems of most concern. The amount and intensity of worry varies greatly across cultures.

Science and technology have in recent years tremendously shrunk the size of the earth in terms of time and space as these philosophical ultimates are related to intercultural contacts. The distance in travel time from San Francisco to Moscow is no greater than was the distance from, say, Chicago to St. Louis only four or five decades ago. The most distant peoples spatially have become near neighbors. Electronics has reduced communication to the farthest reaches of the earth to a matter of minutes; and for neighborhoods to live together peacefully in this rapidly shrinking world they need to understand each other—their cultures, their values, their problems.

The cross-cultural studies of this report may contribute to increased understanding of that segment of the population—the teenagers in school—whose socialization within their respective cultures is going on apace and already largely completed. From among them will come, within a short time, the opinion leaders of their communities.

[1] Given at the Atlantic City meeting of the American Educational Research Association, February 21, 1962.

It is important to note that in the United States, at least, the teenagers have largely interiorized the attitudes and values of the older generation, their parents, who constitute the major portion of the adult population. The teenagers' political preferences quite precisely predict the popular adult vote in presidential elections (Franklin & Remmers, 1960; Remmers, Baker, Fitzhugh, & Heath, 1956; Remmers, Horton, & Mainer, 1952).

Agreement of parents and their teenage children on a variety of attitudes has been shown to be high (Remmers, 1959). Even in matters where literary myth would have violent disagreement between teenagers and their parents—when teenagers should be home from a party, their use of the family car, proper behavior on dates, their own appropriate punishment for misbehavior, and many other issues—careful empirical research has shown much more agreement than disagreement between teenagers and their parents (Hackett, 1951). In brief, the American teenager's mold is apparently largely set. His attitudes or values postdict those of the preceding generation; and one might expect that this generalization is no less and

probably even more true of less open and permissive societies than ours.

SOURCE OF DATA

The studies reported here are based on the findings of four doctoral researches done under the author's supervision and one independent heretofore unpublished study carried out in West Germany. All represent studies by means of translations and adaptations of the Science Research Associates (SRA) Youth Inventory, Form S[2] (de Lopategui, 1957; Kulkarni, 1961; Remmers & Shimberg, 1960; Shimberg, 1949).

Cross-cultural studies present methodological problems of translation and communication. A general assessment of such methodological difficulties is given by Hudson (1959) and others. Restrictions imposed by customs; tabus; values; political considerations; characteristic patterns of interpersonal behavior, including hierarchies of authority and prestige; and the semantic difficulties inherent in translations may all directly or indirectly exert their influence.

The care with which the translations were made is exemplified by de Lopategui (1957). Three independent translations by two psychologists and a high school principal were compared, differences discussed, and a joint agreement reached. The Spanish version was then translated back into English by another psychologist unfamiliar with the original English form. This demonstrated that except for a few minor discrepancies which did not affect the meaning, the Spanish translation was satisfactory.

RESULTS AND DISCUSSION

While the SRA Youth Inventory was orginally primarily designed as an instrument on which to base individual counseling, a purpose the various translations will serve, we are here chiefly concerned with possible generalizations, with uniformities across different cultures. Tables 1 and 2 show that the problems and the cultural values of different cultures to be inferred from them show substantial similarities, thus validating empirically an assertion made nearly 10 years ago (Remmers, 1954):

> Man . . . is, with the exception of a few obvious and largely irrelevant differences in color and size, everywhere the same. Everywhere he faces the same persistent life problems: to gain subsistence, shelter, and security, and to perpetuate the species, to form some sort of group life in which the individual can function in the common affairs of life, to regulate human conduct through the transformation of simple impulsive behavior into behavior patterns consonant with the strivings of the group. This process of socialization or acculturation is perhaps the most distinctively "human" characteristic of the species. Psychologically, the values, the mores, the loyalties, the ideals, or, in short, the attitude patterns thus formed may be thought of as the dynamics of social behavior (pp. 13–14).

The means of Table 1 were computed by weighting the responses to each item 3, 2, and 1 for the large, the middle-sized, the small boxes and 0 for the circle, respectively. This "intensity dimension" is the chief difference between Form S and Form A. On Form A, in contrast, the respondent merely checks any item he perceives as a problem. In view of the fact that the numbers of items in the SRA Youth Inventory differ slightly from one culture to another, these means are only very roughly comparable. That notwithstanding this they show such substantial correlation of

[2] The out-of-pocket costs for these studies were provided by the Purdue Research Foundation from royalties on sales of the Youth Inventory. The author is greatly indebted to E. F. Hylla, H. Roth, and F. Süllwold for the data from West Germany, staff members of the Hochschule für Internationale Paedagogische Forschung, Frankfurt am Main.

TABLE 1

RANK ORDER OF MEAN SCORES OF THE EIGHT PROBLEM AREAS OF THE SRA YOUTH INVENTORY

Problem area	American		Spanish	West German		Indian (Marathi)	
	Boys	Girls		Boys	Girls	Boys	Girls
1. My school	40.99	38.38	26.64	44.01	43.27	63.76	61.80
2. After high school	51.37	50.42	37.13	56.09	55.23	53.80	48.78
3. About myself	34.26	40.72	37.71	34.80	43.28	61.04	61.54
4. Getting along with others	44.85	49.07	41.15	44.99	45.59	66.17	64.55
5. My home and family	27.52	28.89	19.32	33.84	33.33	57.02	50.45
6. Boy meets girl	26.01	28.56	19.74	34.30	34.27	32.68	22.49
7. Health	16.66	20.07	19.28	18.31	18.27	33.09	25.63
8. Things in general	30.58	33.87	34.42	37.04	40.11	67.19	61.37
N	1465	1527	532	283	217	289	81
Rho	.98			.95		.86	

rankings is all the more striking, ranging as they do from .43 between German girls and Indian teenagers to .98 between American youngsters and German girls. The median of these rho's is .74.

Table 3 gives the reliabilities of the eight parts of the inventory and their intercorrelations except for the Spanish version. Unfortunately, the punched cards for it were inadvertently destroyed before the intercorrelations could be computed.

Obviously, the numerical values of the reliability coefficients for each part are highly similar. The psychological similarity of the relative self-perceived importance of these problem areas again is apparent.

TABLE 2

RANK-ORDER CORRELATION OF MEANS OF BOYS AND GIRLS COMBINED

Nationality	Spanish	German boys	German girls	Indian (Marathi)
1. American	71	93	98	66
2. Spanish		79	77	62
3. German boys			95	83
4. German girls				43

Note.—Decimal points omitted.

To probe the meaning of these correlation matrices further factor analyses were computed.[3]

A principle-factor solution was used to obtain the characteristic roots and vectors. The reliabilities for each inventory were used as estimates of the communalities.

In each analysis the first factor accounted for about 65% of the total variance. All of the latent roots were of magnitude 4.9. The factor loadings for the first factor for all the inventories show striking similarity as seen in Table 4.

The second factor accounted for about 9% of the total variance. The magnitude of all these roots was in the neighborhood of .7. Thus the second and third factors were chosen with reluctance. Here the similarities of the factor loadings are obviously rather dissimilar as compared to Factor I, in view of the first factor's heavy loading, accounting for approximately two-thirds of the total variance.

The third factor chosen accounted for approximately 7% of the variance on

[3] These computations were made by the author's assistant, W. H. Payne.

TABLE 3

INTERCORRELATIONS OF THE EIGHT SUBSCORES

Problem area	Subject	1	2	3	4	5	6	7	8
	American	(90)							
	Spanish	(92)							
	German								
1. My school	Boys	(97)							
	Girls	(98)							
	Marathi	(82)							
	American	57	(94)						
	Spanish		(95)						
	German								
2. After high school	Boys	59	(99)						
	Girls	58	(99)						
	Marathi	51	(86)						
	American	63	54	(94)					
	Spanish			(91)					
	German								
3. About myself	Boys	51	57	(98)					
	Girls	62	54	(98)					
	Marathi	64	42	(87)					
	American	57	58	76	(95)				
	Spanish				(95)				
	German								
4. Getting along with others	Boys	64	66	80	(99)				
	Girls	56	57	70	(98)				
	Marathi	57	58	65	(89)				
	American	51	32	68	61	(95)			
	Spanish					(95)			
	German								
5. My home and family	Boys	50	48	68	74	(98)			
	Girls	42	38	46	42	(98)			
	Marathi	57	38	69	58	(91)			
	American	44	48	60	70	55	(94)		
	Spanish						(96)		
	German								
6. Boy meets girl	Boys	48	45	62	67	50	(99)		
	Girls	42	44	42	69	36	(98)		
	Marathi	53	49	51	58	59	(88)		
	American	50	41	65	64	59	68	(91)	
	Spanish							(81)	
	German								
7. Health	Boys	49	47	65	69	63	53	(97)	
	Girls	43	41	66	63	40	49	(95)	
	Marathi	53	32	66	49	66	61	(87)	
	American	38	52	52	62	47	67	71	(96)
	Spanish								(85)
	German								
8. Things in general	Boys	49	61	59	67	53	54	64	(99)
	Girls	39	49	55	55	39	47	47	(98)
	Marathi	43	58	40	66	39	59	39	(90)

Note.—Decimal points omitted. Numbers in parentheses are coefficients of internal consistency.

TABLE 4

UNROTATED FACTOR LOADINGS FOR THE MATRICES SHOWN IN TABLE 3

	United States			India			Germany (Boys)			Germany (Girls)		
	I	II	III	I	II	III	I	II	III	I	II	III
1.	.70	.48	.11	.76	−.08	.12	.73	.43	−.07	.70	.01	.03
2.	.68	.43	−.46	.67	.46	−.14	.74	.53	−.04	.74	−.03	.06
3.	.85	.15	.23	.80	−.30	−.26	.85	−.03	.06	.79	−.44	−.14
4.	.87	.03	.01	.82	.20	−.23	.92	−.01	.03	.86	.02	−.24
5.	.74	−.07	.51	.79	−.35	−.05	.80	−.25	−.25	.63	−.17	.67
6.	.81	−.28	−.11	.78	.05	.41	.74	−.15	.62	.74	.45	−.20
7.	.82	−.29	.02	.74	−.38	.20	.80	−.28	−.19	.73	−.33	−.27
8.	.77	−.36	−.34	.71	.50	.14	.80	.00	−.08	.73	.13	.04

all of the inventories and, again, the similarities of factor loadings were noticed. These roots were of the magnitude .5. Doubtless, error variance would account for at least some of the observed dissimilarities.

In general, then, the factor analysis again supports the similarity of response patterns and their psychological meaning across the cultures at issue.

The averages of Table 1 clearly show considerable variation across cultures. In some cultures youngsters perceive themselves as having more and more important problems than others. To show this in qualitative form Table 5 was prepared. It shows a sample of 20 items with the percentage of "large box" responses for American, German, and Indian boys and girls. In western cultures American youngsters are significantly more worried about learning to read better and about getting people to like them (Items 1 and 118) than youngsters in the other cultures. German teenagers are more concerned about preparing for tests (Item 37), what people do in college (Item 44), what career to pursue (Item 59), being more help to their families (Item 182), and the possible elimination of slums (Item 278).

All of these problems, however, and many more are perceived as "most serious" by a much larger proportion of Indian youngsters. "The heavy and the weary weight of all this unintelligible world" evidently rests most heavily upon them.[4]

In each of the four studies, here reported only in part, many relationships between inventory scores were analyzed—sex, grades, rural versus urban residence, region, socioeconomic status, public versus private schools, caste in India, parents' education, and religion. Space does not permit inclusion of all the many tables of quantitative data. Tables 6 and 7 are abstracted from Kulkarni (1961) and Taliana (1958) to give some indication of these relationships in the two most disparate cultures.[5]

From these tables it is clear that the self-perceived problems are functions of a variety of environmental variables. In Table 6, of the 80 *F* values, 69 are

[4] We have speculated as to whether these results may be a function of response set among Indian students not present at least among American students (Weisbrodt, 1959). S. S. Kulkarni intends to test the hypothesis experimentally.

[5] The interested reader can of course obtain these dissertations either through interlibrary loan or from University Microfilms, 313 North First Street, Ann Arbor, Michigan.

TABLE 5

COMPARATIVE PERCENTAGES OF SOME "MOST SERIOUS PROBLEMS"

Item	American		German		Indian	
	Boys	Girls	Boys	Girls	Boys	Girls
1. I want to learn how to read better.	17	15	6	4	55	53
2. I wish I knew how to study better.	36	36	35	32	68	79
3. I need to learn how to prepare for tests.	4	8	27	28	50	70
19. I wish my teachers would give me encouragement.	9	8	9	10	61	63
35. I wish I understood science better.	19	19	11	17	57	60
37. I worry about getting good grades.	25	23	33	42	58	65
44. I want to know more about what people do in college.	16	21	36	24	60	56
49. I wish I could afford college.	23	21	22	42	53	57
59. What career should I pursue?	19	16	28	32	56	54
81. I feel guilty about things I've done.	12	16	9	11	55	62
118. I want people to like me better.	20	20	5	8	69	65
120. I wish I could carry on a pleasant conversation.	16	15	15	15	69	65
122. I want to make new friends.	24	29	30	25	60	60
123. I need to develop more self-confidence.	22	28	18	28	62	69
142. I need to learn to be a "good sport" in games.	5	6	8	3	60	57
160. It bothers me that some kids are left out of things.	18	27	19	23	50	53
161. I wish I had a quiet place at home where I could study.	12	16	16	11	62	65
182. I wish I could be of more help to my family.	16	17	24	24	65	71
278. Is there any way of eliminating slums?	10	15	22	25	55	58
280. How can I learn to use my leisure time wisely?	12	12	17	18	51	51
$\bar{X}$	16.75	18.00	24.50	21.00	58.80	61.65
N	1,465	1,527	283	217	289	81

TABLE 6

MEAN COMPARISON SIGNIFICANCE TEST VALUES FOR SINGLE CLASSIFICATION ANALYSES OF VARIANCE

Problem area	Rural-urban	Region	Grade	Religion	Caste	Age	Sex	Father's education	Mother's education	Socio-economic status
1. My school	38.5**	41.6**	16.0**	5.0**	9.6**	19.6**	15.6**	5.1**	3.3*	17.3**
2. After high school	26.0**	28.6**	26.4**	14.2**	21.2**	30.6**	68.1**	17.0**	19.5**	63.4**
3. About myself	64.6**	94.8**	15.0**	3.6*	8.8**	24.4**	7.6**	6.4**	2.0	16.2**
4. Getting along with others	28.3**	29.6**	22.9**	7.0**	3.3*	5.5**	30.1**	1.0	2.4	9.3**
5. Home and family	84.8**	165.0**	24.6**	.9	36.7**	42.8**	62.0**	15.5**	12.6**	46.1**
6. Boy meets girl	86.7**	85.0**	.5	7.5**	32.7**	43.1**	94.4**	7.3**	9.4**	22.2**
7. Health	102.9**	81.4**	20.3**	4.7**	33.8**	44.1**	94.6**	16.6**	16.2**	41.1**
8. Things in general	15.0**	27.9**	30.6**	6.7**	1.2	3.1*	51.3**	1.7	2.2	6.9**

Note.—After Kulkarni (1961, pp. 154–162).
* Significant at the .05 level.
** Significant at the .01 level.

significant at or beyond the .01 level, 4 at the .05 level, and only 7 are non-significant. In Table 7, of the 56 values, 44 are significant at or beyond the .01 level, 5 at the .05 level, and 7 are non-significant.

The data presented in this paper support the following conclusions:

1. Teenagers' self-perceived problems can be comparably measured across widely different cultures.

2. The high internal consistency coefficients indicate a reliable measuring instrument.

3. There is a high degree of similarity of rankings of problem areas across cultures.

4. Health problems for teenagers of different cultures are of minimal concern.

5. Post-high-school problems tend to be of major concern.

6. While the relative rankings of problem areas are highly correlated, the amount and intensity of worry vary greatly across cultures.

TABLE 7

MEAN COMPARISON SIGNIFICANCE TEST VALUES FOR SINGLE CLASSIFICATION ANALYSES OF VARIANCE

Problem area	Rural-urban *t* value	Region	Grade	Religion	Sex *t* value	Mother's education	Socio-economic status
1. My school	2.3*	14.4**	12.3**	10.5**	4.3**	19.7**	27.7**
2. After high school	.6	10.0**	9.1**	3.1*	1.2	8.3**	15.6**
3. About myself	2.9**	4.1**	12.9**	5.1**	8.3**	6.9**	10.8**
4. Getting along with others	2.3*	12.1**	10.9**	3.0**	4.7**	.7	10.4**
5. Home and family	2.9**	1.2	7.4**	5.4**	1.6	8.2**	9.0**
6. Boy meets girl	2.7**	2.7*	8.0**	2.7*	2.7**	5.9**	6.7**
7. Health	2.2*	2.2	3.8*	1.8	7.2**	15.5**	13.2**
8. Things in general	4.3**	1.1	1.3	1.3	3.9**	4.5*	12.4**

Note.—Except *t* values as noted. After Taliana (1958, pp. 41–44).
* Significant at the .05 level.
** Significant at the .01 level.

REFERENCES

DE LOPATEGUI, MIGUELINA. Needs and problems of Puerto Rican high school students. Unpublished doctoral dissertation, Purdue University, 1957.

FRANKLIN, R. D., & REMMERS, H. H. Youth looks at civil liberties and the 1960 election. *Purdue Opin. Panel Poll Rep.*, 1960, No. 61.

HACKETT, C. G. Use of an opinion polling technique in a study of parent-child relationships. *Purdue U. Stud. higher Educ.*, 1951, No. 75.

HUDSON, B. B. (Ed.) Problems and methods of cross-cultural research. *J. soc. Issues*, 1959, **15**, 5–20.

KULKARNI, S. S. A Marathi revision of the SRA Youth Inventory. Unpublished doctoral dissertation, Purdue University, 1961.

REMMERS, H. H. *Introduction to opinion and attitude measurement*. New York: Harper, 1954.

REMMERS, H. H. Early socialization of attitudes. In E. Burdick and A. J. Brodbeck (Eds.), *American voting behavior*. Glencoe, Ill.: Free Press, 1959. Ch. 2.

REMMERS, H. H., BAKER, P. E., FITZHUGH, KATHLEEN B., & HEATH, R. W. Musical knowledge and some social attitudes of high school youth. *Purdue Opin. Panel Poll Rep.*, 1956, No. 44.

REMMERS, H. H., HORTON, R. E., & MAINER, R. E. Youth and the 1952 election. *Purdue Opin. Panel Poll Rep.*, 1952, No. 33.

REMMERS, H. H., & SHIMBERG, B. *Manual: SRA Youth Inventory, Form S*. Chicago: Science Research Associates, 1960.

SHIMBERG, B. The development of a needs and problems inventory for high school youth. Unpublished doctoral dissertation, Purdue University, 1949.

TALIANA, L. E. Youth's problems as they see them: A statistical analysis and restandardization of the SRA Youth Inventory. Unpublished doctoral dissertation, Purdue University, 1958.

WEISBRODT, J. A. The effects of response set on the SRA Youth Inventory. Unpublished master's thesis, Purdue University, 1959.

(Received March 5, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 262–270

PARENTAL IDENTIFICATION IN RELATION TO VOCATIONAL INTEREST DEVELOPMENT [1]

JOHN O. CRITES
University of Iowa

The purpose of this study was to derive and test 3 hypotheses about the relationship of parental identification to vocational interest development: (a) degree of identification with the father and mother correlates with vocational interest pattern, (b) kind of identification with the father and mother varies with vocational interest pattern, and (c) pattern of identification is associated with masculinity-femininity of interests. Tests of the hypotheses were made on 3 groups of Ss and Ns equals to 100, 100, and 150. Results from the Semantic Differential and Strong Vocational Interest Blank indicated that a son's identification with both of his parents significantly affects the patterning of his interests, but that his identification with the father is more important.

Several theories of vocational interests suggest that, in addition to intellectual and attitudinal factors, the individual's identifications with significant others, particularly his parents, affect the likes and dislikes he develops. Carter (1940) proposes that identification serves to narrow the range of a person's interests and to focus them in areas of activity and endeavor which lead to the attainment of status and the implementation of the self-concept. Bordin (1943) hypothesizes that the degree of identification with the father directly varies with the relationship of the son's interests to the paternal occupation, the agreement being greater when (*a*) the occupation has high status value, (*b*) the emotional bond between father and son is close, and (*c*) the occupation is traditional in the family. Super (1953), in his theory of vocational development, observes that "childhood and adolescent identifications play a part in shaping vocational interests" (p. 188).

Research on these propositions about identification and interests comes from two different types of investigations: one on the relationship between the interests of fathers and sons, and another on the relationship of independent measures of identification to interests. In two studies of father-son interest similarity, Strong (1943, 1957) obtained an average *r* equal to .30, but with a range which extended from .11 to .48. Berdie (1943) cites unpublished data gathered by Forster which agree with Strong's results, the median coefficient being .33, as well as studies by himself and Dvorak which showed that the sons of skilled workers, businessmen, and physicians have interests like their fathers'. Using *Q* sorts to measure identification, Stewart (1959) found that boys who perceive themselves as their mothers perceive them, both actually and ideally, have more masculine interests. With the same methodology but using college female subjects, White (1959) concluded that girls who describe themselves as both their parents see them have more feminine interests. Finally, Henderson (1958) and Steimel (1960) have reported questionaire findings on high school males which indicate that some relationships exist between identification and (*a*) father-son interest similarity and (*b*) masculinity-femininity of interests.

[1] Presented in abbreviated form at the annual convention of the American Psychological Association, New York, September 2, 1961.

PROBLEM

The results of these studies generally confirm theoretical expectations, but they leave several questions about the role of identification in interest development unanswered: Do individuals who identify more or less closely with their parents have different types of vocational interests? Do individuals who identify more with the father or the mother develop different kinds of vocational interests? And, do individuals who identify with the appropriate parental role model have interests which are consistent with their sex? To answer these questions, it is necessary first to formulate an acceptable definition of *identification*. Although the concept has several connotations (Freud, 1946; Mowrer, 1950; Sanford, 1955), it usually refers to the individual's development of reaction patterns, attitudes, ideals, and values similar to those he perceives in his parents' behavior (Stokes, 1950). As Knight (1940) points out, the process by which a child achieves an identification encompasses, first, a projection of some of his characteristics on to the role model and, second, an introjection of these attributes, along with others idiosyncratic to the model, into his personality (as depicted in Figure 1).

When defined as behaviors which the individual patterns after those he sees

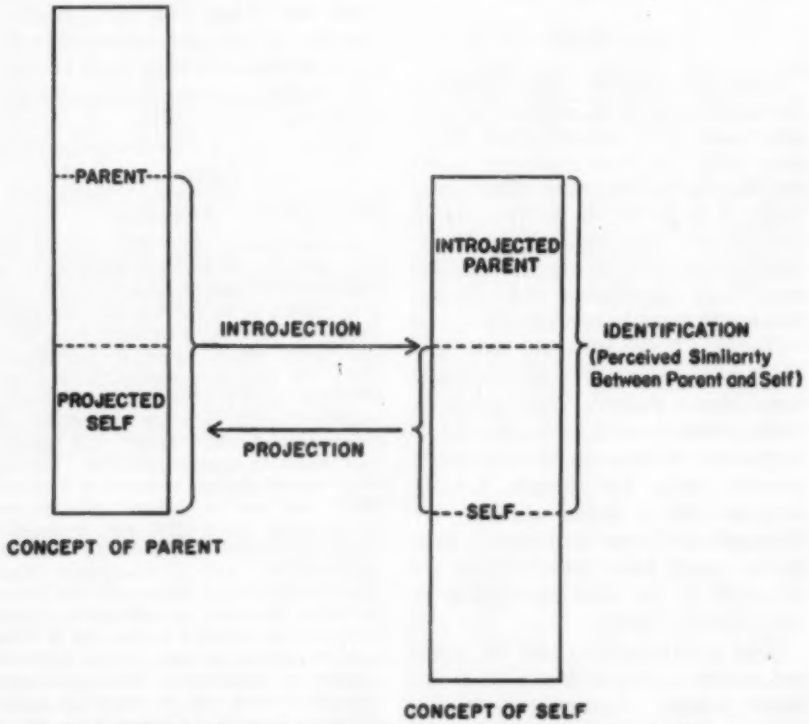

FIG. 1. Identification as the product of projection and introjection of self and parental characteristics (after Knight, 1940).

in his parents, it becomes apparent that identification varies along two different dimensions, one quantitative and the other qualitative. The first, *degree* of identification, expresses the extent of perceived similarity between the individual and his role model: he may identify more or less with his father or mother. The second, *kind* of identification, represents differences in the object with which the individual identifies: he may identify with his father, mother, or both. Together, degree and kind of identification define the *pattern* of identification an individual establishes with his parents: it may be any one of the possible combinations of the two dimensions, such as (*a*) complete identification with the (*b*) cross-sex parent.

HYPOTHESES

From the ways in which parental identification varies in degree, kind, and pattern and from the individual differences which exist in vocational interests, both in type and masculinity-femininity, it is possible to derive a set of three testable hypotheses about identification in relation to interest patterning. These hypotheses, and the rationales for them, are as follows.

Degree of identification with the father and mother correlates with vocational interest pattern. The more an individual tries to emulate the behavior of his parents, the more his interests should resemble theirs. For example, if a son identifies with a father who has the likes and dislikes of an engineer, then the son should have interests which are patterned in the same occupation or field (Bordin, 1943).

Kind of identification with the father and mother varies with vocational interest pattern. According to Strong (1955), some interest areas are more masculine than others, e.g., the technical versus verbal-linguistic occupations. Therefore, an individual who identifies more with one parent than with the other, or who identifies with both parents, should have interests which are consistent with his role models. For example, a mixed-sex identification should produce an interest pattern in occupations, such as those in the social service field, which allow an opportunity for the expression of both masculine (e.g., doing research) and feminine (e.g., helping people) preferences.

Pattern of parental identification is associated with masculinity-femininity of interests. An individual who identifies strongly with either his father or mother should have interests which are appropriately sex-typed. For example, a son who closely follows his mother's example should have interests which are more feminine than those of a son who develops a strong identification with his father.

PROCEDURE

Measurement of Variables

Parental identification. This variable has been defined as the perceived similarity of the individual to his parents, and the best measure of it appears to be the Semantic Differential (SD) (Bieri, Lobeck, & Galinsky, 1959; Lazowick, 1955; Osgood, Suci, & Tannenbaum, 1957). For the measurement of parental identification, the SD consists of nine bipolar, seven-point scales on which the subject rates his concepts of "self," "father," and "mother" with respect to opposed adjectives. The nine scales were selected and validated by Bieri et al. (1959) and are as follows: (1) *potency* (strong-weak, heavy-light, and rugged-delicate); (2) *activity* (fast-slow, hot-cold, and active-passive); and (3) *evaluation* (clean-dirty, fresh-stale, and happy-sad). On the rating forms, the scales are arranged in random order and the sequence of concepts is varied from one page to another, so that systematic response sets are disrupted. Three identification scores are derived from the sum of the squared differences between the ratings for a pair of concepts (Osgood & Suci, 1952): (*a*) F = degree of father identification; (*b*) M = degree

of mother identification; and (*c*) FM = the difference between the F and M scores, or kind of parental identification. Patterns of identification (high F—high M, high F—low M, etc.) are obtained from combinations of F and M scores which fall in the high, average, and low ranges of these two variables, as defined by ±1 standard deviation about the mean.[2]

Vocational Interest Patterning. The Strong Vocational Interest Blank (SVIB) was used to measure interests (Strong, 1943). Patterns of interests were analyzed by a new system proposed by Stephenson (1961), which enumerates all possible combinations of scores on the SVIB profile sheet for primaries, secondaries, and rejects in all 11 interest groups. A primary pattern on the SVIB, for example, consists of a majority of scores in the A and B+ score ranges, where these scores indicate a high degree of similarity between a subject's interests and those of men successfully engaged in different occupations. For a measure of masculinity-femininity of interests, scores on the MF scale of the SVIB were utilized.

Subjects

The SD and SVIB were administered to 350 male subjects subdivided into three groups: The Original group consisted of 100 vocational-educational clients seen at the counseling service of the University of Iowa during the 6-month period from January to June 1960. The Replication group numbered an additional 100 vocational-educational clients counseled during the next 6-month period from July to December 1960. And, the Generalization group included 150 nonclients tested in a large undergraduate psychology course during the winter semester of 1960–61. The characteristics of these samples, with respect to age and intelligence, correspond closely to each other, to the general client and nonclient populations at the University of Iowa (local norms), and to client samples described in other studies (e.g., Merrill & Heathers, 1953). The means, standard deviations, and ranges of the three groups for age (to last birthday) and intelligence (composite percentile rank on the Iowa Scholarship and Placement Tests) are given in Table 1. None of the differences between the group means on these variables was significant at the .01 level.

Method of Analysis

The statistical analyses for each of the samples were the same, the intention being to determine how stable and valid the results were through replication and generalization. Simple analysis of variance was used to compare the subjects with primary and secondary patterns in different interest groups on F, M, and FM (Lindquist, 1953). A two-way analysis of variance was employed to test for trends in the means of the three samples on the dependent variables which were significant in the initial tests (Lindquist, 1947). Other statistical techniques, such as *t* tests and product-moment correlations, were applied as necessary in further analyses of the data. The *N*s for some of the analyses were less than the total samples, because a few subjects had unpatterned interests.

[2] In addition to identification, it is possible to measure both projection and introjection with the SD. The former would be the difference between a subject's perceived similarity with either parent and the parent's self-perception (D − parent's self), and the latter would be the difference between a subject's perceived similarity with either parent and his projection on to the parent [D − (D − parent's self)]. In other words, D is composed of two components, projected self and introjected parent; and these can be defined by the arithmetical operations given above.

TABLE 1

MEANS, STANDARD DEVIATIONS, AND RANGES OF ORIGINAL, REPLICATION, AND GENERALIZATION SAMPLES ON AGE AND INTELLIGENCE

Sample	*N*	Age			Intelligence		
		M	*SD*	Range	*M*	*SD*	Range
Original clients	100	21.14	4.14	17–44	52.08	28.09	4–97
Replication clients	100	20.78	3.11	17–40	52.26	20.31	5–96
Generalization nonclients	150	20.34	2.25	18–35	50.66	26.42	1–99

Note.—None of the differences between the samples in age and intelligence was significant at the .01 level.

TABLE 2

INTERCORRELATIONS, MEANS, AND STANDARD DEVIATIONS OF IDENTIFICATION VARIABLES FOR THE ORIGINAL, REPLICATION, AND GENERALIZATION GROUPS
($Ns = 100$, 100, and 150, respectively)

Variable	Sample	Father	Mother	*M*	*SD*
F	Original			17.32	17.46[a]
	Replication			16.99	14.81
	Generalization			10.63	9.40
M	Original	.41		18.11	14.76
	Replication	.26		18.99	14.37
	Generalization	.25		15.95	13.22
FM	Original	.63	−.42[b]	59.11	17.50
	Replication	.58	−.67[b]	57.67	16.16
	Generalization	.40	−.58[b]	57.58	14.91

Note.—All *r*'s were significant at or beyond the .01 level.
[a] Higher scores indicate less identification with the parents.
[b] The *r*'s for M versus FM are negative because scores for the latter variable are obtained by subtracting M from F. Actually, the correlations between M and FM should be read as positive.

RESULTS

Statistical Assumptions

The assumptions underlying the statistical procedures appeared to be tenable. Edwards' (1954) test for homogeneity of variances indicated some differences between the score dispersions of the interest groups for the dependent variables, but these were largely negligible and nonsignificant. The overall distributions for F and M identification measures were negatively skewed and were consequently converted to *T* scales (McCall, 1939), with the effect of approximating normal distributions within interest groups. The masculinity-femininity scores were normally distributed without transformation.

Identification Variables

Table 2 presents the data on the interrelationships and central tendencies of the identification measures for the three samples. The correlations for any two of the variables, e.g., F versus M or F versus FM, across the three groups were not significantly different at the .01 level, as assessed by *t* tests of extreme values. Similarly, the means were homogeneous, with one notable exception. As tested by simple analysis of variance, the mean father identification scores of the Original, Replication, and Generalization groups were significantly different at the .001 level ($F = 9.78$; $df = 2/347$; $p_{.001} = 6.91$). Separate *t* tests showed that the subjects in the client samples identified considerably less with their fathers than the subjects in the nonclient sample.

Hypothesis 1

Tables 3 and 4 present the findings on the hypothesis that Degree of identification with the father and mother

TABLE 3

ANALYSIS OF VARIANCE FOR VOCATIONAL INTEREST PATTERN AND DEGREE OF IDENTIFICATION WITH FATHER

Sample	*MS*		*df*	*F*
Original clients	Between	141.67	9	2.49*
	Within	56.83	75	
	Total	198.50	84	
Replication clients	Between	128.67	9	2.40*
	Within	53.54	68	
	Total	182.21	77	
Generalization nonclients	Between	264.88	8	6.75**
	Within	39.22	126	
	Total	304.10	134	

* .05 level of significance.
** .01 level of significance.

TABLE 4

ANALYSIS OF VARIANCE FOR VOCATIONAL INTEREST PATTERN AND DEGREE OF IDENTIFICATION WITH MOTHER

Sample		MS	df	F
Original clients	Between	29.67	9	.54
	Within	54.70	75	
	Total	84.37	84	
Replication clients	Between	68.22	9	1.17
	Within	58.12	67	
	Total	126.34	76	
Generalization nonclients	Between	163.13	8	2.37*
	Within	68.78	126	
	Total	231.91	134	

* .05 level of significance.

TABLE 5

ANALYSIS OF VARIANCE FOR VOCATIONAL INTEREST PATTERN AND KIND OF PARENTAL IDENTIFACTION

Sample		MS	df	F
Original clients	Between	321.44	9	2.77**
	Within	116.00	72	
	Total	437.44	81	
Replication clients	Between	343.25	9	2.33*
	Within	147.41	63	
	Total	490.66	72	
Generalization nonclients	Between	240.22	8	2.63*
	Within	91.43	118	
	Total	331.65	126	

* .05 level of significance.
** .01 level of significance.

correlates with vocational interest pattern. They indicate acceptance of the hypothesis only for father identification. Evidently, the father plays a more prominent role in the formation of interest patterns than the mother, at least for males. As depicted in Figure 2, sons who identify closely with their fathers tend to have predominant interests in the business detail occupations, whereas those who identify only slightly with their fathers tend to develop primary interests in the verbal-linguistic field. Most sons identify partially with their fathers, however, and form interest patterns in a variety of occupational groups, including such diverse ones as physical science and social service. These findings are equally valid for the Original,

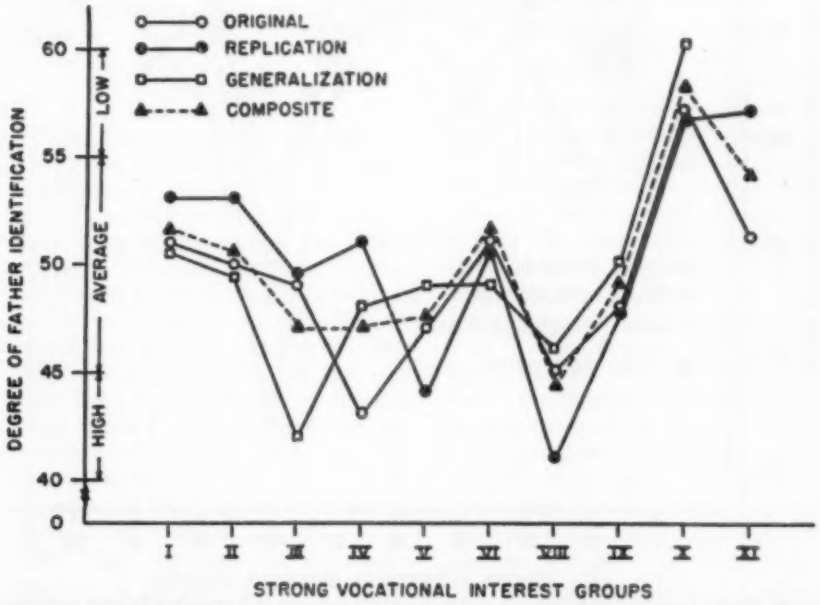

Fig. 2. Father identification means for SVIB primary and secondary interest patterns.

Replication, and Generalization samples, since trend tests revealed no significant differences between them on mean father identification scores for the several interest areas ($F=0.92$; $df=18/256$; $p_{.05}=1.75$).

Hypothesis 2

Table 5 summarizes the results on the hypothesis that Kind of identification with the father and mother varies with vocational interest pattern. The data provide some support for it in each of the samples. As shown in Figure 3, like-sex identifications characterize males who have primary interests in business detail occupations, whereas mixed-sex and cross-sex identifications typify those who have patterned interests in other fields, such as the technical and verbal-linguistic. These tendencies emerge, even though the Degree of identification with the father or mother may be low. In other other words, identifying at least minimally with either one or both parents correlates with the development of an interest pattern. This conclusion applies to all three samples, since trend tests yielded no significant differences between them on FM ($F=0.85$; $df=18/252$; $p_{.05}=1.75$).

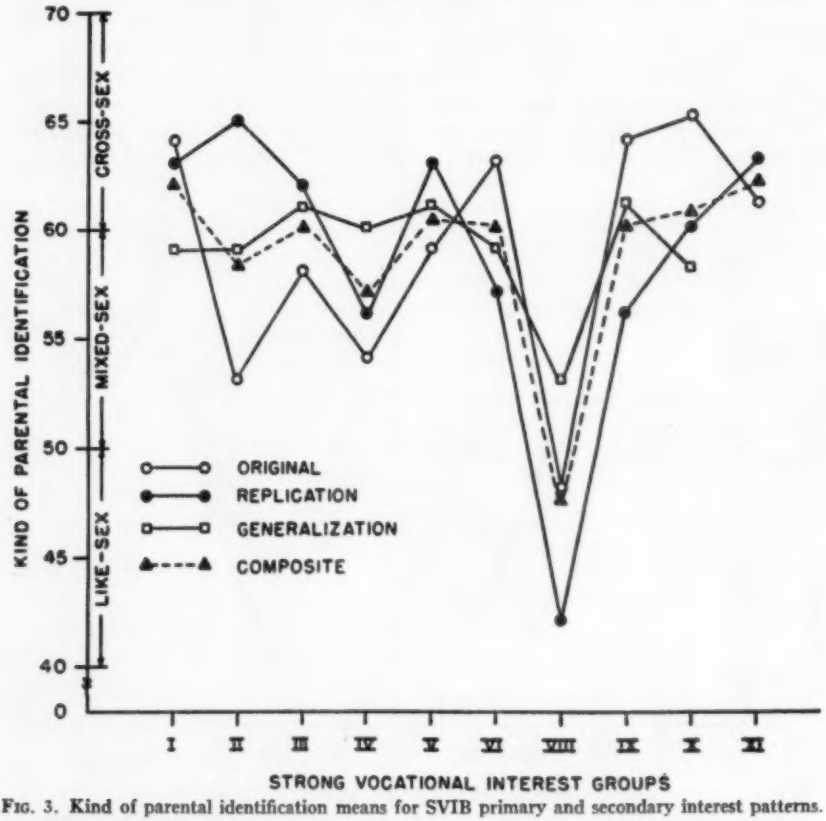

FIG. 3. Kind of parental identification means for SVIB primary and secondary interest patterns.

Hypothesis 3

Finally, the data on the hypothesis that Pattern of parental identification is associated with masculinity-femininity of interests support a significant covariation only in the Generalization group. In this sample, possibly due to its larger N, which makes the statistical test more powerful, there are systematic differences in the mean MF scores for males who identify more or less closely with the appropriate parent as a role model ($F=3.24$; $df=3/16$; $p_{.05}=3.24$). More specifically, close like-sex identifications produce masculine interests, close mixed-sex identifications result in a combination of masculine and feminine interest, close cross-sex identifications yield feminine interests, and negligible identifications with both sexes lead to feminine interests. Stated somewhat differently, the intensity of an individual's identification with either or both of his parents directly influences the masculinity-femininity of his interests.

DISCUSSION

The results on the relationships between the various aspects of parental identification and vocational interest patterning tend to support some of the hypotheses but not others. As predicted (Bordin, 1943), F varies with the kind of vocational interest pattern which an individual develops: sons who perceive themselves as quite similar to their fathers have interests in different areas than those who see themselves as unlike their fathers. In contrast, M appears unrelated to interest patterning: evidently mothers have less impact on the likes and dislikes acquired by their sons than theory suggests (Stewart, 1959). FM, which expresses the individual's relative acceptance of the father or mother as a role model as well as degree of perceived similarity with both parents (FM correlates significantly with F and M in each sample), relates to interest patterns as hypothesized: sons who identify more with one parent than the other have primaries and secondaries in different groups on the SVIB. Finally, pattern of identification may affect the masculinity-femininity of an individual's interests, but the evidence on this relationship is not conclusive, since a systematic trend was found in only one of the samples. Considering the tests of all three hypotheses, the most general conclusion which can be drawn from the results is that for males identification with both parents influences the formation of vocational interest patterns, but identification with the father is more important than with the mother.

The means and intercorrelations of the identification variables, although incidental to the main hypotheses of the study, provide data which reveal an important difference between clients and nonclients. The differences between the mean father identification scores of the client and nonclient groups indicate that males with vocational-educational problems have *not* achieved adequate indentifications with their fathers, a conclusion also supported by no group differences in mean mother identification scores. The questions raised by these findings, which seem reliable and well-substantiated, directly pertain to our traditional conceptualization and conduct of vocational counseling at both the secondary and higher educational levels. For example, how reasonable is it to expect a client to develop an appropriate masculine identification as a basis for the resolution of his vocational problems within the 2 or 3 interview hours of typical vocational counseling? How effective are the usual techniques of test interpretation and presentation of occupational information in the facilita-

tion of the identification process? And, what are the implications of the male client's underidentification with a father figure for the sex of his counselor? These and related questions need serious consideration, if we are to adapt vocational counseling to what is known about the nature of vocational problems.

Finally, the interrelationships of the identification measures are quite similar to those obtained in previous studies (Bieri et al., 1959; Payne & Mussen, 1956; Sopchak, 1952). There are low positive correlations between degrees of father and mother identification and moderate positive correlations of each with kind of identification. In other words, the young adult male identifies with general familial patterns of attitudes and behavior as well as the distinctive attributes and characteristics of each parent, much as Freud (1927) observed in his early writings on ego and super-ego development (pp. 40–44).

REFERENCES

BERDIE, R. F. Factors associated with vocational interests. *J. educ. Psychol.*, 1943, **34**, 257–277.

BIERI, J., LOBECK, R., & GALINSKY, M. D. A comparison of direct, indirect, and fantasy measures of identification. *J. abnorm. soc. Psychol.*, 1959, **58**, 253–258.

BORDIN, E. S. A theory of vocational interests as dynamic phenomena. *Educ. psychol. Measmt.*, 1943, **3**, 49–65.

CARTER, H. D. The development of vocational attitudes. *J. consult. Psychol.*, 1940, **4**, 185–191.

EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.

FREUD, ANNA. *The ego and the mechanisms of defense*. New York: International Universities Press, 1946.

FREUD, S. *The ego and the id*. London: Hogarth, 1927.

HENDERSON, H. L. The relationships between interests of fathers and sons and sons' identification with fathers. Unpublished doctoral dissertation, Teachers College, Columbia University, 1958.

KNIGHT, R. P. Introjection, projection, and identification. *Psychoanal. Quart.*, 1940, **9**, 334–341.

LAZOWICK, L. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, **51**, 175–183.

LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.

MCCALL, W. A. *Measurement*. New York: Macmillan, 1939.

MERRILL, R. M., & HEATHERS, LOUISE B. A comparison of the Wechsler-Bellevue and ACE tests on a university counseling center group. *J. consult. Psychol.*, 1953, **17**, 63–66.

MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.

OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean differences and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

PAYNE, D., & MUSSEN, P. H. Parent-child relations and father identification among adolescent boys. *J. abnorm. soc. Psychol.*, 1956, **52**, 358–362.

SANFORD, R. N. The dynamics of identification. *Psychol. Rev.*, 1955, **62**, 106–118.

SOPCHAK, A. L. Parental "identification" and "tendency toward disorders" as measured by the MMPI. *J. abnorm. soc. Psychol.*, 1952, **47**, 159–165.

STEIMEL, R. J. Childhood experiences and masculinity-femininity scores. *J. counsel. Psychol.*, 1960, **7**, 212–217.

STEPHENSON, R. R. A new pattern analysis technique for the SVIB. *J. counsel. Psychol.*, 1961, **8**, 355–362.

STEWART, L. H. Mother-son identification and vocational interest. *Genet. psychol. Monogr.*, 1959, **60**, 31–63.

STOKES, S. M. An inquiry into the concept of identification. *J. genet. Psychol.*, 1950, **76**, 163–189.

STRONG, E. K., JR. *Vocational interests of men and women*. Palo Alto: Stanford Univer. Press, 1943.

STRONG, E. K., JR. *Vocational interests 18 years after college*. Minneapolis: Univer. Minnesota Press, 1955.

STRONG, E. K., JR. Interests of fathers and sons. *J. appl. Psychol.*, 1957, **41**, 284–292.

SUPER, D. E. A theory of vocational development. *Amer. Psychologist*, 1953, **8**, 185–190.

WHITE, BECKY J. The relationship of self concept and parental identification to women's vocational interests. *J. counsel. Psychol.*, 1959, **6**, 202–206.

(Received March 26, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 271–274

RESPONSES OF BRIGHT, NORMAL, AND EMH CHILDREN TO AN ORALLY-ADMINISTERED CHILDREN'S MANIFEST ANXIETY SCALE[1]

NEIL A. CARRIER, KENNETH D. ORTON
Southern Illinois University
AND LESLIE F. MALPASS
University of South Florida

Groups of children 10–14 years old differentiated (by WISC IQ) into bright (mean IQ 129), normal (mean IQ 102), noninstitutionalized educable mentally handicapped (EMH) groups (mean IQ 67), and an institutionalized EMH group (mean IQ 67) were given the Children's Manifest Anxiety scale (CMA scale) orally and individually. CMA scale scores generally differentiated the groups, and girls tended to have higher A scale and L scale scores than boys. Tentative evidence of a negative A scale IQ relationship as well as a possible acquiescence set in low IQ Ss suggest caution in using the CMA scale.

The number of recent studies employing the Children's Manifest Anxiety (*CMA*) scale indicates the popularity of this revision for children (Castaneda, McCandless, & Palermo, 1956) of the Taylor Manifest Anxiety (*MA*) scale (Taylor, 1953). However, the studies reported do not always show consistent findings. While normative and relational data on this instrument have been reported by a number of investigators (Castaneda, McCandless, & Palermo, 1956; McCandless & Castaneda, 1956; McCandless, Castaneda, & Palermo, 1956; Malpass, Mark, & Palermo, 1960; Palermo, 1959), differences between studies in the age and intelligence range of subjects tested, the method of administering the scale, and the kinds of other measures employed have probably contributed to the discrepancies between results.

A previous study (Malpass et al., 1960) compared the scores of groups of institutionalized and public school educable mentally handicapped (EMH) children and normal children on the *CMA* scale. But, whereas most studies have administered the *CMA* scale as a pencil-and-paper group test, Malpass et al. (1960) met the problem of reading deficiency in the retarded subjects by using an oral-individual administration procedure in testing them, but used the usual procedure in testing normal subjects.

There are three major features of the present study: (*a*) the use of subjects selected and classified by the Wechsler Intelligence Scale for Children (WISC) into bright, normal, and EMH groups (the last category subdivided into institutionalized and noninstitutionalized subjects), thus enabling a determination of whether the *CMA* scale differentiates these groups; (*b*) the use of the oral-individual method of administering the *CMA* scale with all subjects, thereby avoiding the interpretation problem imposed by two different administration procedures that the Malpass et al. (1960) study encountered, and also providing additional normative data on this method of administration;

[1] This study is part of research project No. 578 (SAE 8182) sponsored by the United States Office of Education Cooperative Research Program and by the Graduate School, Southern Illinois University.

(*c*) the use of a slightly older group of children than has generally been used.

METHOD

Subjects

Children ranging in age from 11 to 14 years from five southern Illinois communities were classified by WISC IQ into bright, normal, and noninstitutionalized educable mentally handicapped (NIEMH) subjects. The 32 bright subjects (IQ interval 120–150) had a mean age of 12-7 years and a mean IQ of 129. The 31 normal subjects (IQ interval 90–110) had a mean age of 12-7 years and a mean IQ of 102. The 27 NIEMH subjects (IQ interval 50–80) were selected from five public school special classes. Their mean age was 12-11 years and mean IQ was 67. A group of 17 institutionalized educable mentally handicapped (IEMH) subjects were tested at the Lincoln State School and Colony in Illinois. The interval of this group was also 50–80; the mean age was 12-3 years and the mean IQ was 67.[2]

Procedures

An oral-individual procedure for administering the *CMA* scale was employed with all subjects. The experimenter carefully read each of the 53 statements to each subject recording each "yes" or "no" response as it was given.[3] A previous check had found that EMH subjects attend to this procedure and apparently comprehend the items (Malpass et al., 1960).

Employing simple analysis of variance, each of the two scales—anxiety (*A* scale) and lie (*L* scale)—of the *CMA* scale was used as a criterion measure with overall group differences determined separately by sex. Significance of differences between adjacent classification groups were then computed using *t* ratios for independent samples. Zero-order correlations between *A* scale scores and both IQ and CA were obtained for each group and sex within each group.

[2] Grateful acknowledgment for assistance in the data collection phase of this study is made to school officials in Carbondale, Du Quoin, Herrin, Marion, Murphysboro, and West Frankfort, Illinois, and to J. Albaum, Superintendent, and W. Chambers, Chief Psychologist, Lincoln State School and Colony.

[3] Three of the IEMH subjects gave "yes" responses to every statement in the scale; one gave only one "no" response. Their data were discarded.

RESULTS AND DISCUSSION

The results as presented in Table 1 show that overall group differences were significant at the .001 level for both sexes on both scales. In Table 2 it can be seen that the *A* scale significantly differentiated the classification groups excepting on the normal-NIEMH comparison.

There was some confirmation of sex differences in *A* scale scores as has been found in several studies (e.g., Castaneda, McCandless, & Palermo, 1956). However, the only such sex difference significant at the .05 level was found for NIEMH subjects. While this is in apparent agreement with a similar finding in the Malpass, Mark, and Palermo (1960) study, problems in obtaining an appropriate NIEMH sample from a restricted population resulted in a discrepancy of 14 IQ points between sexes (means: boys, 72; girls, 58) which could account for this result.

Compared to the study most similar to the present one in terms of kind of subjects used (Malpass et al., 1960), the means for the classification groups on the *A* scale show an irregular tendency to be somewhat higher. In the earlier study "normal" subjects (actually including both normal and bright subjects as defined in the present study) scored a mean of 13.7. Combining our bright and normal groups produces a fig-

TABLE 1

ANALYSIS OF VARIANCE OF CLASSIFICATION GROUPS

Source	*A* scale *df*	*A* scale *MS*	*A* scale *F*	*L* scale *df*	*L* scale *MS*	*L* scale *F*
Boys						
Groups	3	664.31	12.38*	3	57.25	14.24*
Error	64	53.64		64	4.02	
Total	67			67		
Girls						
Groups	3	469.79	11.48*	3	51.87	10.44*
Error	35	40.91		35	4.97	
Total	38			38		

* Significant at the .001 level.

TABLE 2

DIFFERENCES IN *A* SCALE AND *L* SCALE SCORES BETWEEN SELECTED CLASSIFICATION GROUPS

Group	*N*	A Scale			L Scale		
		M	*SD*	*t* ratio	*M*	*SD*	*t* ratio
Bright, all	31	12.93	5.10		2.45	2.01	
				3.69*			.70
Normal, all	32	18.71	7.01		2.84	2.30	
				.57			3.88***
NIEMH, all	27	19.51	5.16		5.18	2.19	
				6.15***			3.44**
IEMH, all	17	30.64	6.56		7.17	1.24	
Bright boys	20	12.65	4.41		2.50	1.96	
				2.75**			−.19
Normal boys	21	17.71	6.66		2.38	1.85	
				.15			3.80**
NIEMH boys	18	18.00	4.55		5.05	2.32	
				4.85**			1.35
IEMH boys	9	30.56	8.26		6.55	1.17	
Bright girls	11	13.45	6.14		2.36	2.10	
				2.40*			1.24
Normal girls	11	20.64	7.25		3.73	2.80	
				.59			1.49
NIEMH girls	9	22.56	5.98		5.54	1.89	
				3.15**			3.15**
IEMH girls	8	30.75	3.83		7.87	.93	

Note.—*t* ratios are for adjacent groups.
* $p \leq .05$.
** $p \leq .01$.
*** $p \leq .001$.

ure for comparison of 15.8. The IEMH subjects in the present study had a mean *A* scale score of 30.6, while those in the previous study had a mean *A* scale score of 25.9. However, the NIEMH subjects in the present study scored lower than those in the earlier one, 19.5–21.4. These differences may be due to chance variation, to the fact that our subjects averaged about a year older than subjects previously used, and/or to the influence of the oral-individual administration procedure. That the latter may produce some (probably small) elevation in *A* scores over those obtained by the usual procedure is not unlikely. This should be tested in a study specifically designed for this purpose.

Significant differences between the classification groups on the *L* scale (Table 2) were found between the normal and NIEMH groups both for boys and for sexes combined, and between the EMH groups both for girls and for sexes combined. Not shown is the significant sex difference for IEMH subjects. In three of four comparisons a tendency is shown for the girls to have higher scores than boys. This is consistent with previous findings (Malpass et al., 1960).[4]

Within-group correlations between *A* scale scores and IQ were computed. The correlation for bright subjects was −.01, for normal subjects, −.35 (significant .05 level), for all EMH subjects combined, −.31 (significant .05 level), for NIEMH subjects, −.32, and for IEMH subjects, −.40. The all-

[4] B. R. McCandless, personal communication, 1961.

EMH-boys correlation of −.41 was also significant at the .05 level. Although based on samples of restricted variance, it is interesting to note that with but one exception (bright girls) all of the 15 *A* scale—IQ correlations by sex and within-group were negative.

When all classification groups were combined the correlation between the *A* scale and IQ was −.54 (.01 level); with the bright, normal, and NIEMH groups combined it was −.39 (.01 level); with the bright and normal groups combined it was −.45 (.01 level). It should be noted, however, that by thus combining groups more subjects are included in the tails of the distribution than would likely be found in a random sample of the population. The correlations presented are probably thereby inflated.

The general impression from the correlations in the previous two paragraphs, however, is that control on IQ may be desirable in studies relating *A* scale scores to other variables. For example, several investigators have used the *CMA* scale as a measure in studies relating anxiety (or "drive") and complex learning (Castaneda, Palermo, & McCandless, 1956; Palermo, Castaneda, & McCandless, 1956). If *A* scale scores are indeed negatively related to IQ, it is possible that the poorer performance of high anxiety subjects observed on the more complex tasks could be partially explained as a function of intelligence.

Of interest is the finding that IEMH subjects as a group (and by sex) gave more responses representing anxiety than did the NIEMH subjects. This finding is in general agreement with the Malpass, Mark, and Palermo (1960) study. But experience with the testing of these groups suggests care in interpreting this finding. The investigators gained a definite impression of an "acquiescence set," in questioning EMH subjects, which was particularly noticeable with the institutionalized subjects. Although neither the observation of such a tendency nor speculation about its likely instrumental qualities in the life of the retarded person are original with the writers, a caution is thereby raised against naively assuming differences in "anxiety" or "drive level" between the groups solely on the basis of *A* scale differences.

In summary, it appears that the *CMA* scale administered orally and individually generally differentiated the classification groups of this study. A tendency was found for girls to have higher *A* scale scores and *L* scale scores than boys. But the negative correlations of IQ with *A* scale scores, and the suspicision of an acquiescence set in the EMH subject caution in inferring the bases upon which the observed differentiation was effected.

REFERENCES

CASTANEDA, A., MCCANDLESS, B. R., & PALERMO, D. S. The children's form of the Manifest Anxiety scale. *Child Develpm.*, 1956, **27**, 317–326.

CASTANEDA, A., PALERMO, D. S., & MCCANDLESS, B. R. Complex learning and performance as a function of anxiety in children and task difficulty. *Child Develpm.*, 1956, **27**, 327–332.

MCCANDLESS, B. R., & CASTANEDA, A. Anxiety in children, school achievement, and intelligence. *Child Develpm.*, 1956, **27**, 379–382.

MCCANDLESS, B. R., CASTANEDA, A., & PALERMO, D. S. Anxiety in children and social status. *Child Develpm.*, 1956, **27**, 383–391.

MALPASS, L. F., MARK, S., & PALERMO, D. S. Responses of retarded children to the Children's Manifest Anxiety Scale. *J. educ. Psychol.*, 1960, **51**, 305–308.

PALERMO, D. S. Racial comparisons and additional normative data on the Children's Manifest Anxiety Scale. *Child Develpm.*, 1959, **30**, 53–57.

PALERMO, D. S., CASTANEDA, A., & MCCANDLESS, B. R. The relationship of anxiety in children to performance in a complex learning task. *Child Develpm.*, 1956, **27**, 333–337.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received April 9, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 275–278

CHANGES IN STUDENT TEACHERS' ATTITUDES TOWARD CHILDHOOD BEHAVIOR PROBLEMS

GEORGE A. PINCKNEY
University of Nebraska

To determine the extent to which a course influences attitudes, 203 students in introductory educational psychology ranked 32 childhood behavior traits on the basis of seriousness. Rankings were made during the 1st and last week of classes, and changes in rank order were analyzed and compared with those of a control group of peers and with the rankings of clinicians in a study done a decade earlier. Precourse rankings of the control and experimental groups were not reliably correlated with the 1951 clinicians. On postcourse rankings the correlation between the control group and clinicians remained insignificant while the correlation between the clinicians and the experimental group was highly reliable, suggesting that courses can change attitudes.

Educators have long been aware of the difficulty teachers have in judging pupil behavior in its true psychological perspective. Teachers have a tendency to overrate the seriousness of behavior problems related to sex, dishonesty, and disobedience, while underrating shyness, sensitiveness, fearfulness, and other behavioral manifestations which do not express pupil antagonism toward authority in the classroom.

Wickman (1928) found that teacher ratings of the seriousness of 50 behavior problems correlated negatively with ratings given by 30 clinicians. Rank-order correlation between the ratings of 28 teachers in Cleveland and the clinicians was found to be −.22, while a group of 511 Minneapolis teachers correlated −.11 with the same 30 clinicians.

This, and similar studies (Amos & Washington, 1960; Gage & Suci, 1951; Gronlund, 1950; Hunter, 1957; Thompson, 1940), have drawn attention to the need for more understanding and education on the part of student teachers concerning the behavior problems of young children. Teacher training institutions have expended considerable quantities of both time and money investigating the most suitable coursework and supervised experience programs to include in their respective curricula.

Twenty years after Wickman's study, Schrupp and Gjerde (1953) replicated the Wickman rating procedure to determine what, if any, progress had been made in teachers' ability to interpret psychological problems manifested in child behavior. They found that differences were still present in teacher-clinician ratings, but that the magnitude of these differences was not as great as it had been in the earlier study.

This closer agreement between the classroom teacher and the clinician can be assumed to have been brought about, in part at least, by the general increase in public knowledge of basic psychological concepts, as well as by the adoption of curriculum policies which require one or more courses in psychology in teacher education programs.

The present study was designed to investigate the extent to which changes in rating of childhood behavior problems are related to the experience in psychology courses for teachers college students.

Method

Subjects

The experimental group for this study consisted of 203 undergraduate students enrolled

in six sections of an introductory educational psychology course in Teachers College, University of Nebraska. The control group was composed of 58 students enrolled in the same college, but who had not taken coursework in any area of psychology. All of the subjects in both the experimental and the control group were planning for careers in the teaching profession.

Procedure

At the first-meeting of each section of the educational psychology courses the subjects were presented with a list of 32 behavior traits selected from Wickman (1928). Instructions were given to rank these traits from the most serious to the least serious using a nine-point *Q* sort technique after Stephenson (1953). A list of the behavior traits under consideration in the present study may be seen in Table 1.

Subsequent course material was presented in the usual manner for the remainder of the semester, with no further mention of the ranking being made. All reference to the Wickman (1928) study was carefully removed from lecture notes, and reading assignments were

TABLE 1

RANKING OF 32 BEHAVIOR TRAITS

Traits	Experimental preranking	Experimental postranking	Control preranking	Control postranking	1928 teachers	1928 clinicians	1951 clinicians
Stealing	1 (6.48)	5 (6.12)	1 (6.67)	2 (5.94)	2	11.5	11
Cheating	2 (5.52)	14 (5.14)	2 (6.06)	3 (5.87)	7	18.5	20
Destructiveness	3 (5.29)	3 (6.45)	6 (5.35)	5 (5.53)	8	29	14
Untruthfulness	4 (5.28)	8 (5.73)	3 (5.43)	4 (5.59)	5	18.5	19
Cruelty	5 (5.24)	1 (7.45)	4 (5.41)	1 (6.84)	6	6	3
Heterosexual activity	6 (5.11)	9.5(5.50)	16 (3.87)	6 (5.25)	1	20	21
Disobedience	7 (4.97)	16 (5.08)	5 (5.38)	7 (4.84)	9	26.5	24
Obscene notes	8 (4.96)	22 (4.70)	12 (4.61)	12.5(4.34)	4	22.5	25
Temper tantrums	9 (4.94)	6 (6.06)	8 (5.12)	16 (3.84)	10	15	10
Profanity	10 (4.94)	23 (4.43)	7 (5.29)	8 (4.65)	11	30	30
Masturbation	11 (4.62)	18 (5.00)	11 (4.64)	15 (3.93)	3	26.5	28
Selfishness	12 (4.56)	15 (5.09)	9 (4.83)	14 (4.03)	19	14	22
Impudence	13 (4.33)	21 (4.75)	13 (4.38)	9 (4.59)	12.5	24	27
Unhappiness	14 (4.15)	2 (6.80)	10 (4.68)	23 (3.30)	17.5	3	2
Overcritical of others	15 (4.10)	17 (5.05)	23 (3.19)	11 (4.47)	29	8	9
Resentfulness	16 (4.07)	7 (5.92)	14 (4.22)	10 (4.50)	20	4	13
Unsocialness	17 (4.03)	4 (6.26)	18 (3.64)	17 (3.72)	26	1	1
Smoking	18 (3.76)	30 (2.85)	27 (2.90)	29 (2.62)	14	32	31
Domineering	19 (3.76)	19 (4.92)	20 (3.51)	12.5(4.34)	23	10	17
Suspiciousness	20 (3.68)	13 (5.17)	19 (3.55)	18 (3.66)	25	2	5
Easily discouraged	21 (3.56)	9.5(5.50)	17 (3.68)	19 (3.62)	17.5	7	7
Laziness	22 (3.39)	27 (3.79)	15 (3.90)	24 (3.22)	12.5	25	23
Tattling	23 (3.33)	28 (3.34)	22 (3.29)	21 (3.47)	30	22.5	29
Enuresis	24 (3.13)	25 (4.21)	24 (3.16)	26.5(2.90)	15	21	16
Nervousness	25 (3.09)	12 (5.30)	25 (3.03)	28 (2.87)	16	16.5	6
Cowardliness	26 (2.98)	24 (4.24)	21 (3.42)	20 (3.61)	22	13	18
Fearfulness	27 (2.93)	11 (5.34)	30 (2.01)	25 (2.97)	24	5	4
Interrupting	28 (2.84)	32 (2.61)	26 (2.97)	22 (3.37)	28	31	32
Tardiness	29 (2.71)	31 (2.68)	28 (2.80)	26.5(2.90)	21	28	26
Sensitiveness	30 (2.62)	20 (4.91)	29 (2.42)	31.5(2.03)	31	9	15
Shyness	31 (1.85)	26 (3.81)	31 (1.61)	30 (2.34)	32	11.5	8
Dreaminess	32 (1.53)	29 (3.19)	32 (1.32)	31.5(2.03)	27	16.5	12

Note.—Mean scores on the *Q* sort are shown in parentheses.

screened to avoid, as much as possible, any reference to it or similar work.

The subjects in the control group were chosen at random from sections of courses in history and principles of education, and were administered the *Q* sort with the same instructions as the experimental group.

The same procedures were followed with both the experimental and control groups during the final week of the school year. The *Q* sorts were scored on a nine-point basis, with the trait ranked as least serious receiving no points. A rank order of trait seriousness based upon mean *Q* sort scores was established for precourse ratings by both the experimental and the control groups, and a correlational analysis of these was made.

RESULTS

The rankings of the 32 behavior traits, together with mean rating for the experimental and control groups, are shown in Table 1. The results of the correlational analysis of the precourse and postcourse rankings for both the experimental and the control groups are shown in Table 2. A ranking of the 32 behavior traits based on the earlier work of Wickman (1928) and Gage and Suci (1951) was made for the 1928 teachers, 1928 clinicians, and the 1951 clinicians. These results are also included in Table 2.

The precourse ranking of the control group was found to correlate .939 with the experimental groups' precourse ranking of the behavior traits. The precourse ranking of the control group correlated .854 with their own ranking of the traits at the end of the course. The postcourse ranking of the experimental group, however, correlated only .555 with the control groups' postcourse ranking and .608 with their own ranking made prior to the course.

The control group correlated −.152 with the 1951 clinicians on the precourse rankings and −.029 with the postcourse ranking. The experimental group, on the other hand, correlated −.069 with the 1951 clinicians on the precourse ranking and .663 on the postcourse ranking.

The significance of the change of the experimental group from a correlation of −.069 with the 1951 clinicians in the precourse ranking to a correlation of .663 on the postcourse ranking was determined by transforming the rho's to *Z* values and expressing the differences as a relative deviate of the normal curve as suggested by Wert, Neidt, and Ahmann (1954). This change was found to be significant at the .001 level of confidence. The same change for the control group was found to be nonsignificant.

TABLE 2

INTERCORRELATIONS AMONG GROUP RATINGS ON 36 BEHAVIOR TRAITS

Group	1928 clinician	1951 clinician	1961 control (precourse)	1961 control (postcourse)	1961 experimental (precourse)	1961 experimental (postcourse)
1928 teachers	−.340	−.220	.764**	.650**	.816**	.394*
1928 clinicians		.860**	−.063	−.017	−.107	.557**
1951 clinicians			−.152	−.029	−.069	.663**
1961 control (precourse)				.854**	.939**	.603**
1961 control (postcourse)					.901**	.555**
1961 experimental (precourse)						.608**

* Significant at .05 level.
** Significant at .01 level.

DISCUSSION

It is apparent from Table 2 that the attitudes of the experimental group toward childhood behavior problems changed to a significant degree. A correlation of −.069 was found between this group and the 1951 clinicians on the precourse rankings. The rankings at the end of the course correlated .663, indicating a change was significant at the .001 level of confidence. No such change was apparent in reference to the control group. The precourse and postcourse correlations of the controls with the 1951 clinicians were −.152 and −.029, respectively.

The data would tend to substantiate the hypothesis that psychology courses may play a significant role in the formulation of teachers' attitudes toward childhood behavior problems. An integral part of this educative process must certainly be the attitudes which the instructor himself feels are important. These may be expressed formally in class lectures or may be interpreted by students from the less overt behavior patterns of the instructor. Their relative contribution to the total change in attitudes is not known, but it must be assumed that they exert a larger influence on the experimental group than on the controls. This must be kept in mind when interpreting the data of the experimental group.

It may be noted that of the five traits rated as most serious by the 1951 clinicians (unsocialness, unhappiness, cruelty, fearfulness, and suspiciousness) four were among the traits registering the greatest change from precourse to postcourse rankings of the experimental group. The greatest disagreement between the clinicians and the experimental group of the postcourse ranking was found on the traits of shyness, dreaminess, and heterosexual activity.

In conclusion, it seems apparent from the results of this investigation that introductory courses in psychology can meet the needs of student teachers for better understanding of the basic principles of behavior.

REFERENCES

AMOS, R. T., & WASHINGTON, R. M. A comparison of pupil and teacher perceptions of pupil problems. *J. educ. Psychol.*, 1960, **51**, 255–258.

GAGE, N. L., & SUCI, G. Social Perception and teacher-pupil relations. *J. educ. Psychol.*, 1951, **42**, 144–152.

GRONLUND, N. E. The accuracy of teachers' judgments concerning the sociometric status of sixth grade pupils. *Sociometry*, 1950, **13**, 197–225, 329–357.

HUNTER, E. C., Changes in teachers' attitudes toward children's behavior over the last thirty years. *Ment. Hyg., N. Y.*, 1957, **41**, 3–11.

SCHRUPP, M. H., & GJERDE, C. M. Teacher growth in attitudes toward behavior problems of children. *J. educ. Psychol.*, 1953, **44**, 203–214.

STEPHENSON, W. *The study of behavior.* Chicago: Univer. Chicago Press, 1953.

THOMPSON, C. E. The attitude of various groups toward behavior problems of children. *J. abnorm. soc. Psychol.*, 1940, **35**, 120–125.

WERT, J. E., NEIDT, C. O., & AHMANN, J. S. *Statistical methods in educational and psychological research.* New York: Appleton-Century-Crofts, 1954.

WICKMAN, E. K. *Children's behavior and teachers' attitudes.* New York: Commonwealth Fund, 1928.

(Received April 19, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 279–287

DISCERNMENT OF TEACHER CHARACTERISTICS BY TAT SEQUENCE ANALYSIS

M. INNOCENTIA BURKARD
Mount Mary College

What are some of the underlying personality factors that distinguish teachers rated by their pupils at the extremes in teaching efficiency? 300 religious women teaching in Grades 4 through 12 were rated by their classes (10,720 pupils). TAT Sequence Analysis of 1200 stories of high-rated and low-rated paired samples yielded clearly distinguishing characteristics in the 2 groups ($p < .001$). The high-rated teachers showed a basic disposition to respond to life realistically and constructively, while the low-rated ones did not. Results were verified by the successful assignment of the teachers to the high or low group on the basis of their positive or negative scores on the empirically established criteria.

While many investigators have found differences in the observable behavior of teachers who were rated at the extremes in teaching efficiency, not many have studied the underlying aspects of the teacher's personality that might account for the differences in overt behavior (Barr & Jones, 1958). The present investigation was based on the hypothesis that such fundamental differences between teachers who are rated high by their pupils and those who are rated low can be reached by a disguised technique, specifically, a projective test. For this purpose the Thematic Apperception Test (TAT) was selected as the most promising for its penetration of the deeper motivational levels of the personality; and Arnold's TAT Sequence Analysis (Arnold, 1949, 1960) was judged the most objective method of analyzing and scoring the results.

METHOD

Criterion Groups

The criterion groups were set up on the basis of pupil ratings according to the Diagnostic Teacher-Rating Scale by Amatora (1950). The Scale consists of an Area scale calling for a rating on a five-point continuum in each of seven divisions: liking for teacher (I); teacher's ability to explain (II); kindness, friendliness, and understanding (III); fairness in grading (IV); discipline (V); amount of work required (VI); liking for lessons (VII). Following this is a Diagnostic Check List consisting of 49 scaled statements. The latter are arranged in 7 intrascales of 7 items for each of the above divisions. These items are in turn scaled so that their values extend from highest to lowest, with the fourth item in each intrascale at the midpoint.

The original sample of this investigation consisted of 300 teachers, all of whom were religious women teaching in parochial and private elementary and secondary schools in the Midwest. Fourteen high schools and 23 grade schools were represented. The teachers ranged in age from 20 to 68 years, and from 1 to 48 years in teaching experience. The schools varied in size, the number of faculty members ranging from 1 to 40. In every case but two, the entire faculty was included in the rating.

The raters were a total of 10,720 pupils, 6,226 in Grades 4–8, and 4,494 in Grades 9–12. They included both boys and girls, and, with the exception of chance absentees, the entire class of each teacher rated.

The rating scale was administered by the investigator to each class in its own homeroom. Pupils were not identified, and teachers only by code number. As a means of preparing the pupils, the second part of the scale, the Diagnostic Check List of 49 items, was administered first. The pupils were asked to check each item that was true of this teacher. On the Area scale, which followed immediately, the pupils were asked not to *rate* their teacher as called for on the scale but to *rank* her relative to all the other teachers they had

ever had. The numbers thus indicated rank positions and could be considered arbitrary weights or scores for purposes of calculation; for example, each teacher who was ranked at the top of a particular pupil's teachers received a score of 5 for the area on which she was so ranked.

Each teacher's mean score on each of the seven areas was computed, using the designated figure in the five-point scale as a score. Inspection revealed that these means became progressively lower from the fourth grade to the twelfth, indicating that the pupils became more critical and discriminating as they grew older. To have a common basis for comparison, therefore, centile norms were set up in each of the seven areas for each of the following grade levels: 4, 5–6, 7–8, 9–10, 11–12 (Burkard, 1958; Innocentia, 1962).

Out of the original 300 teachers a sample was made consisting of 25 pairs of elementary teachers and 25 pairs of secondary teachers, matched on age and intelligence. In the pairing process, such factors as size and location of school, grade level, and subject matter taught were also controlled as far as possible. The high member of each pair had received rank scores in the upper centiles on five out of the seven areas of the rating scale, while the low member was at the opposite extreme in the same five areas. For pairing purposes, Divisions IV and VI, which correlated least with the rest of the scale and which were difficult to interpret because of varying regulations with regard to grading and assignments, were ignored. The median rank score of the elementary teachers on the remaining five areas was 86.7 for the high group and 16.3 for the low; in the secondary sections it was 85.5 for the high group and 23.9 for the low. Table 1 presents the comparative statistics of the criterion groups on all the variables.

TABLE 1

COMPARATIVE STATISTICS OF THE PAIRED GROUPS

Variables		Elementary teachers		Secondary teachers	
		High rank	Low rank	High rank	Low rank
	N	25	25	25	25
Age	Mean	37.7	38.7	42.6	43.6
Experience [a]	Mean	17.3	17.9	20.8	19.6
IQ [b]	Mean	113.9	113.5	113.9	116.2
Rank score [c]	Median	86.7	16.3	85.5	23.9
TAT score [d]	Mean	10.2	1.9	10.4	1.8

[a] Years of teaching experience in private or parochial schools.
[b] Gamma IQ on Otis Intelligence Test.
[c] Centiles based on 5 rank scores.
[d] Number of positive statements as measured by the Scoring Standard.

Measures

The TAT was administered to the 300 teachers in groups ranging from 2 to 23, each teacher writing 12 stories. Eleven pictures and the blank card were used in the following order: 1, 2, 3BM, 4, 5, 6BM, 7BM, 8BM, 9BM, 10, 16, 11. The standard directions were given, with emphasis on writing a story that contained a plot and an outcome. There was no time limit; the cards were numbered and displayed about the room; and the stories were written in an unbroken period that varied according to individual needs, from 55 to 95 minutes. In order to produce a comparable measure for the control of intelligence, each teacher, at another session, did Form Am, Gamma, of the Otis Quick-Scoring Mental Ability Test.

TAT Analysis

The method of TAT analysis used in this study is less generally known than others and, therefore, requires some elaboration. Arnold's Sequence Analysis has developed from the observation that a series of stories gives a sequential account of the writer's life situation. This is observable when the series is considered from the viewpoint of the "import" of the stories taken in sequence as written. By the *import* (Arnold, 1960) is meant a brief condensation of the moral of a story without interpretation (p. 342). It is a restatement in a generalized, abstract form of what the writer is saying by means of a particular, concrete situation. This may be illustrated by the following story taken from a TAT protocol.

> 1. Johnny pleaded with his father to take lessons on the violin, but when the teacher prescribed an hour's practice daily, he found it hard to give up his play time. A little sulkily he sat making a decision. With a sigh for the fun his buddies were having, he picked up his violin and began his scales. Father was reading the newspaper in his armchair.

In the above story the writer is "actually" saying:

> You may want something; but then, when you find out it entails work and sacrifice,

you may not want to stick to it; and you do so only because you can't get out of it.

It is characteristic of these generalized statements always to follow a sequence. If the abstraction is correctly done, there will appear an association among these statements from story to story; and this feature makes of the total protocol a more or less continuous, connected expression of the subject's way of looking at his world. A series of stories taken from the same protocol and succeeding the example above, illustrate this sequence (see Table 2).

When used in a clinical situation the Sequence Analysis reveals the client's area of preoccupation or conflict. But the analysis also reveals how the client is handling his conflict, since each statement expresses an attitude that may be designated as either constructive or not constructive. In the above series for instance, Import 4 may be considered more constructive than any of the others.

In the present research it was this latter aspect of the analysis, that is, the nature of the expressed attitudes, that was studied. No prior judgment was made regarding the "con-

TAT PROTOCOL WITH SEQUENCE ANALYSIS

TABLE 2

Story as written	Import
2. In early spring Laura's father always had to work doubly hard, and her mother lost some of her beauty with each passing year. But Laura just couldn't let all their troubles come between her and her books. She just wouldn't end up a farmer's daughter, she thought as she turned away.	2. (But) you won't let anything, not even the troubles of your loved ones, come between you and what you want most.
3. Six yards of sheer organdy from Aunt Bess! This was what had assured Jean that she might accept John's invitation to the prom. Coming home from school happily dreaming of pretty ruffles and soft music, Jean had found little Janice with mother's scissors, and her dreams destroyed. She knelt at the window seat and wept bitterly. She knew her answer must be, "No."	3. (And) when someone destroys your dream, and you know you cannot have it, you weep bitterly.
4. "Don't go," Mary pleaded as she caught Joe's shoulder in a wild effort to restrain him. Then she glanced around their home, for it was home in spite of the cheap prints and gaudy furnishings. Joe signed up though, because he too loved and he was a man.	4. (But) if you are strong, you can give up what you have for a greater love.
5. "So she put them there again," Mrs. Carol noted with some surprise. Every day the girl cut flowers and placed them on Mr. Jack's study table. "I ought to stop these going-on," she muttered. But Mr. Jack paid cash and on time. This is what stopped her.	5. You may know what you ought to do, but you don't do it because you fear to lose by it.
6. His mother stood with her back to him as he poured out his story. Why didn't she turn and tell him he was a no-good son, gambling away his life and hers, wasting his wages, leaving her nothing but her widow's pension to eke out her existence. But she wouldn't. She turned and put her hand on his shoulder with a smile.	6. You know you are no good and admit it, but still they do not reproach you.
7. A dream come true! The old father smiled a contented man's smile as his boy poured out his determination to give up that alluring woman and to finish his engineering course. One more year and he would have a stalwart partner.	7. They keep on hoping in you as you tell of your determination to give up what lures you away and to persevere.

Table 2—Continued

Story as written	Import
8. It was by pure accident that the rifle shot had struck old Mike in the stomach. Albert had grown into a man as they stretched Mike out for the emergency operation. Only after they assured him that Mike would be all right did he turn away. There was a dream in his eyes and a determination. He, too, would save lives with his intelligence and steady hand.	8. It may have been a tragic accident that brought you to that determination.
9. He raised his head and studied the sleeping faces around him. There was Tom sprawled out with his head on Joe's stomach. Sam lay snoring. Everyone was relaxed; but he was tense, excited with his first taste of battle. In just one hour it would begin again.	9. (And) the first taste of the struggle may have excited you more than others.
10. Now they had been married forty years and Jim was just going off to work as usual. Mary smiled as she said, "God bless your day." Jim picked up his lunch pail and set out with the same old song in his heart.	10. (But) after many years you are still content to do the same things in the usual way.
11. The boy stood with his hand on the railing, looking up at the door with the little rivers of rain running down its shiny varnish. With his left hand he reached in to feel the warm report card stiff against his heart. Father was home and there would be no rewards this time.	11. A day of reckoning will come and it will bring no reward.
12. The wounded lamb had to be left on the narrow path just before the men had ascended the rude bridge. A vulture circled over it menacingly. As the last man disappeared in the valley, the bird darted down and plunged its beak into the soft flesh.	12. (And) when you are left behind wounded and unprotected, someone will be ready to strike at you.

structiveness" of the statements, the object being to find out what kind of attitudes were expressed predominantly by each criterion group. If the imports of the groups differed to a degree that they successfully discriminated between the high and the low teachers when the latter were not identified as such, these imports would in effect express the differences in the two groups on the level of motivation that the TAT explored.

ANALYSIS AND RESULTS

Construction of the Scoring Standard

The TATs of the 100 cases in the paired sample were analyzed according to Arnold's method of TAT Sequence Analysis described above. The results were in the form of 12 sequential statements (1 for each of 12 stories) for each case. The protocols of 10 pairs of teachers were selected at random from the secondary group for the purpose of setting up a "scoring key" empirically on the basis of the statements of these 20 teachers. Such a scoring key, if it successfully discriminated between the high and low teachers when they were not identified as such, would in effect express the differences in the two groups on the level of motivation that the TAT explored. For this purpose the TATs of the 10 selected pairs were scrutinized for types of statements that appeared predominantly among the high teachers and not among the low, and vice versa. Statements that appeared only occasionally in either group and expressed a basic disposition contrary to the predominant one were transferred to the opposite set. Two opposing classes of statements were thus built up, one of which represented the dispositions expressed by the high group, and

the other, those expressed by the low group. The former were designated *plus;* the latter, *minus*.

Inspection revealed that both plus and minus statements readily fell into five categories having to do with the following aspects of life: success, failure, loss (including disappointments, injustice, and hardships), the obligations of life, and other people. Each of the assembled statements was assigned to the category to which it had reference.

A trial test was made to determine whether these sets of statements could be used with the elementary teachers to distinguish the groups according to their rank. Ten pairs, identified only by code number, were selected at random from among the elementary teachers. The cases were shuffled, and the investigator tried to predict blindly, on the basis of the categorized statements alone, which teachers had been rated high and which low. This attempt demonstrated that the compiled statements were not sufficiently inclusive for a workable standard. These elementary cases were, therefore, studied for differentiating characteristics in the same way the first had been. The results were found to be of the same general nature as those of the secondary group; and they were combined with the latter, making the "key" more descriptive and, hence, easier to use with other protocols. Table 3 presents illustrative statements from the Scoring Standard, somewhat generalized, though following closely the phraseology of the Sequence Analysis.

Application of the Standard

The remaining 60 cases were then evaluated according to the Scoring Standard. The cases were not identified as to pairs nor rating. Each statement in each subject's sequence analysis was marked either plus or minus according to the standard; and, on the basis of the total positive statements, each case was assigned to the high or the low group. Two errors were made in this attempt: one high-rated case was inaccurately judged low, and one low case, high.

To test the results further the 100 TAT protocols and the Scoring Standard were given to two other judges, doctoral students in psychology who had been trained in sequence analysis. The judges scored the protocols independently, and both accurately placed every teacher in the high or low group on the basis of her plus and minus scores. On the 1,200 stories taken as a whole, there was 97.2% and 96.6% agreement between the investigator and Judges A and B, respectively. Agreement between Judges A and B was 94.3%.

Final scores were then assigned to all cases. The scores were based on the decisions of two out of the three judges—that is, if two judges agreed that a statement had plus value, the statement was given a positive score. Allowing 1 score point to each statement, the positive scores of the high teachers ranged from 8 to 12, with a mean of 10.2 in the elementary group and 10.4 in the secondary. The range of positive scores in the low groups was from 0 to 4, the mean being 1.9 for the elementary and 1.8 for the secondary (Table 1).

For a measure of statistical significance of the differences between the high and the low teachers, the sign test (Siegel, 1956) was applied to the pairs of scores. Since all the differences were in the same direction, that is, in every pair of teachers the high-rated one received more plus scores, x (the number of fewer signs) was 0. For $N=25$, an x of this size is significant beyond the .001 level of confidence.

TABLE 3

ILLUSTRATIVE STATEMENTS FROM THE SCORING STANDARD

Plus	Minus
Categories A and B: Success and Failure	
Achievement depends on constructive action: effort, work, sacrifice, preparation; realistic adaptation; supernatural or ethical motives; help or advice from others. Resolutions to achieve require recognition of difficulties involved. Failure results from inadequacy in oneself; it can be overcome by constructive action.	Success comes in spite of uncertainty, compromise. Success is only a dream. The goal is achieved easily, unrealistically, without constructive action. Resolutions are made impulsively, without foresight or planning. Failure just happens, or is due to others. One is dejected or seeks to escape in dreams or in easy work. Failure is not admitted, or is overcome without constructive effort. Failure ends in despair or frustration.
Category C: Loss, Harm, Danger	
Something is actively done about a loss. Loss may be overcome by constructive action. Negative emotions are overcome. Loss is accepted when required by the nature of things or for the good of oneself or others or by one's duty. There is hope in spite of loss.	Loss is due to others and causes negative emotions. Loss is not accepted; it always remains a tragedy. Loss is accepted passively without effort to overcome it. Loss is averted by others or by fate. Loss leads to impulsive or desperate action; it ends in despair or destruction.
Category D: Life and Its Obligations	
Legitimate enjoyment succeeds; ill-intentioned action is punished. Revenge is rejected, frustrated, punished. Duties are undertaken for rational reasons; they are the right thing to do; they help others. Right conduct is rewarded; love is a help in doing right.	Duties are undertaken and decisions made because there is no way out, or just to please. Right conduct brings failure, loss; love makes life easier. The view of life is unrealistic or illogical: life is all work, worry, emptiness; life is all sunshine. Work is drudgery. Wrongdoing is condoned, or is regretted for any reason other than conviction of its wrongness.
Category E: People	
Others are good, responsive, helpful. Good relationships are attained by cooperative action: good-will, accepting one's limitations realistically, avoiding or escaping bad companions.	Others are frustrating, unforgiving, unhelpful. One gives in to others even though they are not right. One depends on others entirely. Good relationships are attained out of fear or not at all. Bad relationships lead to no real problems.

To determine which areas of the rating scale were most significant in identifying the teachers with plus characteristics, phi coefficients were computed between the teachers' scores on the TAT and their rank scores on each of the seven areas. The two categories of TAT scores were the number of teachers making scores of 0–4 and the number making scores of 8–12. (There were no scores between 4 and 8.) The rank scores were dichotomized at the fiftieth centile. Table 4 shows these correlations. The coefficients ranged from .37

TABLE 4

RELATIONSHIPS (PHI COEFFICIENTS[a]) BETWEEN TAT SCORES AND SCORES ON THE SEVEN AREAS OF THE RATING SCALE

TAT score	Areas of rating scale						
	I	II	III	IV	V	VI	VII
Elementary ($N = 50$)	.96	.85	.92	.69	.88	.68	.96
Secondary ($N = 50$)	.96	.81	.77	.41	.73	.37	.78

Note.—All coefficients significant at .01 level.
[a] Phi coefficients are restricted in size and are, therefore, lower than Pearson r's. The phi coefficients in this study are further reduced by the use of Yates' correction for continuity. These coefficients may, accordingly, be looked upon as conservative estimates of the correlation between the given variables.

to .96. Lowest relationships were with Areas IV (grading) and VI (assignments) on both levels; highest, with Areas I (liking for the teacher) and VII (liking for the lesson) in the grade school, and with Area I in the high school.

DISCUSSION

The study supports the hypothesis that underlying personality differences between teachers ranked high by their pupils and those ranked low can be found by TAT Sequence Analysis. While the teachers were ranked on the basis of their external classroom behavior, the TAT revealed their motivational pattern in the form of convictions or attitudes. These convictions or attitudes are remarkably different for the two groups.

The complete Scoring Standard, derived as it is directly from the TAT protocols, delineates these differences in detail. With regard to achievement, for instance, there is a strong contrast between what may be designated the active attitude of the high group and the passive attitude of the low. The teachers in the high group constantly reiterate the need of action on their part to achieve success or to overcome failure. They foresee difficulties and recognize the need of planning and preparation. For them, success follows on such activity. The low group, on the other hand, do not mention these requirements of success. According to their stories, success just comes easily or in some equally unrealistic way; or, again, it does not come at all in spite of constructive activity on the achiever's part. This group generally attributes failure to extraneous causes, rather than to inadequacies in themselves (Table 3, Categories A and B).

The same active and passive attitudes are apparent toward loss and hardship. Here the activity of the high group is directed toward overcoming the loss, or, if that is impossible, toward bearing it without self-pity, resentment, or other negative emotion. The low-ranked teachers appear overcome by losses, and either express futile hostility or grief, or remain inert (Category C).

With regard to life and its obligations, the high group reveals an objective, principled view. Right is accepted just because it is right; and what is seen as good conduct is rewarded, while wrongdoing fails or is punished. The reverse is expressed in the imports of the low group. To these teachers life appears all drudgery or all joy with no adequate reason for either. Again, their passive attitudes are apparent in an inability to do anything toward bringing about that joy or eliminating the drudgery (Category D).

Toward other people the high-ranked group expresses a willingness to go along reasonably, seeing them as good and helpful. The low group, on the contrary, sees people either as hostile or as friendly in a sentimental way. Again, adequate reasons for the hostility or friendliness are not apparent. Good relationships are seen as unattainable, or

if attained, as leading to no good results (Category E).

The fact that these attitudes were given in a projective test suggests that they are habitual ways of thinking and feeling of which the owner is for the most part only dimly aware. Because of their habitual nature, the convictions are undoubtedly strongly motivational, inducing actions that are consonant with them. Further evidence of this is the high correlation between the TAT scores and rank scores, suggesting that the attitudes expressed in the TAT result in actions that account for the corresponding ranking.

If this is true, the teacher ranked high would appear to be an active person, a hard worker, who recognizes her own limitations and also the demands of achievement; a person who takes an objective, realistic view of things and whose judgment is not clouded by negative emotions. Her classroom behavior is probably similarly objective, well organized, and efficient.

The low-ranked person appears to be the reverse of all these. She expects success without working for it; and, when faced with failure or difficulty, acts in a less reasonable way. Apparently not seeing the relationship between the events of life and her own initiation of, or reaction to them, she tries to explain life in terms of accident or the ill will of others. Carried over to the classroom such attitudes probably result in poorly planned lessons, inefficient procedures, and emotional reactions to the consequent failures.

Several unanswered questions present themselves. Do the expressed attitudes denote several characteristics or could they all be reduced to, let us say, rationality versus emotionality? Is the unrealistic way of seeing things just one more effect of an inability to direct emotion reasonably? Or is there a cognitive factor operating here in addition to emotional control? This area needs to be further explored.

The scoring system used in this study was very simple, merely giving one point for each plus statement, and not allowing for differences in the degree to which the positive or negative attitude is expressed. A 4-point scoring system, ranging from $+2$ to -2, has been devised since this study was completed, providing more flexibility in scoring. Further work needs to be done on this.

It may be of interest to point out that the criterion in this study was not teaching efficiency as such, but rather the pupils' rating of teaching efficiency. Whether these are the same thing or not has never been demonstrated empirically. Investigators of teaching efficiency have labored under the difficulty of finding a valid, reliable criterion. The objective here was to locate differences in the teachers as the latter were evaluated by their pupils. The results seem to indicate that the religious woman who is ranked high by her pupils is a high-caliber person as far as the dimensions measured by TAT Sequence Analysis are concerned.

Again, are the pupils merely reacting to a "well-liked" person when they rank a teacher high? From Table 4 it is apparent that the highest correlations with TAT scores are to be found in Area I, "liking for the teacher." Area II, having to do with how clearly the teaching is done, also correlates high with the scores; but how much of this is halo effect? On the other hand, the pupils' rating is based, not on samples of the teacher's classroom behavior, but on all of it. This fact, coupled with a consideration of the calibre of the high-ranked and the low-ranked teacher, suggests that the pupils' rating might be as true an estimate of the teacher's efficiency and personality as the ratings of

part-time observers or the indications of other criteria.

In spite of these unanswered questions the study reveals the divergent ways of thinking of the two criterion groups. The fact that the subjects were all religious women limits the degree to which the scoring criteria can be applied to teachers in general. It is reasonable to propose, however, that the technique of TAT analysis used here, since it succeeded so well in distinguishing high and low groups, could well be used to set up scoring criteria for other teachers and eventually to predict teaching success.

REFERENCES

AMATORA, S. M. A diagnostic teacher-rating scale. *J. Psychol.*, 1950, **30**, 395–399.

ARNOLD, M. B. A demonstration analysis of the Thematic Apperception Test in a clinical setting. *J. abnorm. soc. Psychol.*, 1949, **44**, 97–111.

ARNOLD, M. B. *Emotion and personality*. Vol. 2. New York: Columbia Univer. Press, 1960.

BARR, A. S., & JONES, R. E. The measurement and prediction of teacher efficiency. *Rev. educ. Res.*, 1958, **28**, 258–264.

BURKARD, M. I. Characteristic differences, determined by TAT sequential analysis, between teachers rated by their pupils at the extremes in teaching efficiency. Unpublished doctoral dissertation, Loyola University, Chicago, 1958.

INNOCENTIA, S. M. What do pupils demand of their teachers? *Cath. educ. Rev.*, 1962, **60**, 179–184.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956. Pp. 68–75.

(Received April 23, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 288–292

THE RELIABILITY OF MEASURES OF HANDWRITING QUALITY

LEONARD S. FELDT
State University of Iowa

This paper reports a reliability analysis of measures of handwriting quality. Components of variance techniques are employed to assess the magnitude of measurement errors arising from various sources. The data indicate that the reliability of quality measures, as typically obtained in elementary school classrooms, is quite low. In Grades 1 and 2 the coefficients for measures based on a single day's writing sample range from .4 to .6. Systematic variation among teachers in the harshness of their judgments is shown to be a particularly important source of error. Suggestions are offered for improving the reliability of handwriting evaluations.

Scales for evaluating the quality of pupil handwriting were among the first standardized measuring instruments to be used in American schools. Thorndike's Scale for Handwriting of Children was published in 1910, several years before the first intelligence tests found their way into the classroom. Ayers' (1912) A Scale for Measuring the Quality of Handwriting of School Children, which is still widely used, and Freeman's (1915) An Analytical Scale for Judging Handwriting appeared soon after the Thorndike scale in the period 1912–15.

Although handwriting scales have been in use more than 50 years, relatively little research has been reported regarding their reliability. Only one of the commercially distributed scales was backed at publication by reliability data, and virtually no studies have subsequently appeared which deal specifically with this problem. The major purpose of this investigation was to gather such data for one of the more recently published handwriting scales.

Experimental Design

The present study was concerned with the reliability of the Evaluation Scales, Grades 1 and 2, which accompany *Guiding Growth in Handwriting* (Freeman, 1959). These two scales are used to evaluate manuscript writing and include specimens representative of five levels of quality. Instructions for using the scale indicate that pupil writing samples may be assigned scores higher than 90, the value for the best of the scale specimens, or lower than 65, the value for the poorest of the scale specimens. These directions were followed in carrying out the study.

Subjects

The subjects who took part included 30 first grade pupils and 29 second grade pupils enrolled in the University Elementary School of the State University of Iowa. This school uses the series *Guiding Growth in Handwriting* (Freeman, 1959). The pupil population of the University Elementary School is considerably above average in intelligence and achievement. Many of the children are sons and daughters of university faculty members and probably receive more parental encouragement in their school work than do typical public school children. However, their achievement in handwriting is not above average. The data reported below do not reflect the superiority which generally characterizes their average achievement in other school subjects.

Procedure

Each child contributed three samples of writing, following the procedures specified in the teacher's edition of *Guiding Growth in Hand-*

writing. The first sample was obtained on December 8, the second on December 13, and the third on December 16, 1960. The separation of testing dates was considered adequate to permit study of day-to-day fluctuations in quality. It was sufficiently short, on the other hand, to rule out any real improvement in overall writing quality.

The three judges who evaluated the writing samples were drawn from the enrollees in an Iowa University course in kindergarten and primary teaching methods. Two of the judges were graduate students and one was a senior undergraduate. All of the judges had had some teaching experience in the lower elementary grades. Preliminary investigation suggested that the judges needed supervised practice in the use of the Evaluation scales. Three practice sessions were held, in which samples of handwriting were projected on a screen via an opaque projector. The judges viewed the samples, made a tentative evaluation of quality, compared ratings, and then discussed their differences. The purpose of these sessions was to sharpen perceptions of overall quality and to develop consistent standards regarding the emphasis to be given the individual aspects of slant, spacing, clarity of letter form and letter size.

In the experiment proper, the three judges evaluated the test samples independently. The papers were identified only by pupil number, and submitted to each judge in a different random order. Each judge recorded his ratings on a separate score sheet seen only by himself.

The statistical method used to analyze the data was that known as the components of variance technique. This methodology yields not only reliability coefficients but also independent estimates of the error variance arising from various sources. Five categories of measurement error were hypothesized for ratings of handwriting quality: (*a*) momentary variations in the quality of pupil writing, as reflected within the sample taken in a single testing session; (*b*) undependability or inconsistency arising from the judging process; (*c*) day-to-day fluctuations in the quality of the writing performance; (*d*) systematic variation among judges in the emphasis given various aspects of quality; and (*e*) systematic variation among judges in the overall leniency with which they judge writing samples. The experimental design used in this study yielded separate estimates of the last three components of error variance and a combined estimate of the first two. While independent estimates of the first two components could have been obtained, the additional labor this would have required of the judges did not seem justified.

RESULTS

The analyses of variance for Grades 1 and 2 are presented in Tables 1 and 2; the estimates of the several variance components for each grade are given in Table 3. The types of factors represented by each component are summarized below.

σ^2_e: Variation in handwriting quality within a sample taken on a given day. Also reflected is error of measurement arising from subjectivity of the judgment process.

$\sigma^2_{D\times S}$: Day-to-day fluctuations by individual subjects, beyond that attributable to σ^2_e. (This factor is referred to as day $\times$ subject, $D \times S$, interaction.)

$\sigma^2_{J\times S}$: Inconsistencies among judges in their ranking of all pairs of subjects, beyond that attributable to σ^2_e. This component reflects inconsistencies arising from different subjective weightings of the elements of quality. (This factor is referred to as judge $\times$ subject, $J \times S$, interaction.)

σ^2_J: Variations in harshness from one judge to another.

σ^2_S: True variation among experimental subjects; score variation arising from sources other than errors of measurement.

TABLE 1

ANALYSIS OF VARIANCE OF QUALITY SCORES OF GRADE 1

Source	*df*	*MS*	*F*
Subject (*S*)	29	290.4425	
Day (D)[a]	2	1.7175	.10
Judge (J)[b]	2	155.1575	17.11*
D × J[c]	4	5.3875	1.23
D × *S*	58	16.4600	
J × *S*	58	9.0675	
D × J × *S*	116	4.3700	

[a] $F_D = MS_D/MS_{D\times S}$.
[b] $F_J = MS_J/MS_{J\times S}$.
[c] $F_{D\times J} = MS_{D\times J}/MS_{D\times J\times S}$.
* $p < .01$.

TABLE 2

ANALYSIS OF VARIANCE OF QUALITY SCORES OF GRADE 2

Source	*df*	*MS*	*F*
Subject (*S*)	28	200.9175	
Day (D)[a]	2	28.5175	1.29
Judge (J)[b]	2	165.2600	13.86*
D × J[c]	4	9.7575	1.42
D × *S*	56	22.1600	
J × *S*	56	11.9275	
D × J × *S*	112	6.8800	

[a] $F_D = MS_D/MS_{D\times S}$.
[b] $F_J = MS_J/MS_{J\times S}$.
[c] $F_{D\times J} = MS_{D\times J}/MS_{D\times J\times S}$.
* $p < .01$.

TABLE 3

ESTIMATES OF VARIANCE COMPONENTS FOR QUALITY SCORES: GRADES 1 AND 2

Component	Estimate	
	Grade 1	Grade 2
1. σ^2_e	4.3700	6.8800
2. $\sigma^2_{D\times S}$	4.0300	5.0925
3. $\sigma^2_{J\times S}$	1.5650	1.6825
4. σ^2_J	16.1200	16.7175
5. σ^2_S	29.9200	19.3000
Standard error of measurement (Components 1–3)	3.1567	3.6952

A comparison of the four error variance components reveals that judge-to-judge differences in overall harshness or leniency, as reflected in σ^2_J, is by far the most serious. This is true in both grades studied. The error variance arising from the measuring procedure itself, σ^2_e, is not extremely large, considering the magnitude of true subject differences, σ^2_S. Day-to-day fluctuations by individual subjects, $\sigma^2_{D\times S}$, represent about as potent a source of error as the within-day, within-judge component. The judge-to-judge inconsistencies, on the other hand, were rather small. The size of $\sigma^2_{J\times S}$, relative to σ^2_e, suggests that judges were weighting the several aspects of the writing in much the same way in arriving at their judgments. The error components for Grades 1 and 2 are fairly consistent in size. It appears that judges encountered no greater difficulty in applying one grade scale than the other. There did appear to be greater heterogeneity in writing quality within the first grade population than within the second grade population. This inference can be drawn from the comparative size of σ^2_S for the two grade groups.

A number of reliability coefficients may be derived from these components, just as there are a number of correlation approaches by which test reliability may be approximated. The essential differences among the coefficients derived by either method arise from differences in the handling of the individual error components. For example, one might use two judges to score a single set of papers and correlate the results. Or one might have the first judge score writing samples collected on Tuesday and the second judge score a new set of samples collected from the same children on Thursday. The correlation between the latter pairs of judgments would almost certainly be lower than the first correlation, since the day-to-day fluctuations of the pupils

are reflected in the second coefficient and not in the first. A number of reliability coefficients for each grade are reported in Table 4. The coefficients were derived using formulas presented in Lindquist (1953). To clarify the differences among these coefficients, each is described in terms of the hypothetical correlation coefficient to which it corresponds. It may be noted that several of these coefficients involve reuse of the same writing samples or repeated judgments by the same judge. Such coefficients are based on a restricted definition of errors of measurement. Measurement authorities agree that the preferred reliability coefficient is that value which reflects the measurement error arising from *all* sources. According to this principle, a coefficient such as the first one suggested above would be unsatisfactory. It tends to exaggerate the reliability of the evaluation process, since it ignores sources of error which have actually affected the children's scores. If all error sources are taken into account, the fourth coefficient represents, for a single measure, the most realistic appraisal of reliability.

TABLE 4

RELIABILITY ESTIMATES FOR GRADES 1 AND 2

Correlation counterpart	Value for current study	
	Grade 1	Grade 2
Scores assigned by Judge A correlated with scores assigned by Judge B for the same writing samples.	.851	.740
Scores assigned by Judge A on Sample 1 correlated with scores assigned by Judge A on Sample 2 taken several days later.	.789	.637
Scores assigned by Judge A on Sample 1 correlated with scores assigned by Judge B on Sample 2 taken several days later.	.750	.586
Scores assigned by Judge A on Sample 1 correlated with those assigned by Judge B on Sample 2 taken several days later; judge differences considered as error of measurement.	.534	.389
The mean of two scores (per pupil) taken on Day 1 and evaluated by Judge A correlated with the mean of two scores taken on Day 2 and evaluated by Judge B.	.794	.654
The mean of three scores (per pupil) taken on Days 1, 2, and 3 and evaluated by Judge A correlated with the mean of three scores taken on Days 4, 5, and 6 and evaluated by Judge B.	.873	.773

DISCUSSION

By present-day achievement test standards, the reliability coefficients reported in Table 4 are rather low. Those for Grade 2 are particularly disappointing. The standard error of measurement, even when variation in judge harshness is excluded, amounts to over half the difference between the scores of adjacent scale specimens. However, these data should not be taken as evidence that handwriting scales do not serve a useful purpose. Handwriting constitutes an important area in the elementary curriculum, and scales of this type appear to provide the best means for handwriting evaluation. Thus, one might look for ways in which reliability can be raised, rather than rejecting the scale for its weaknesses. Two avenues can be suggested by which reliability may be improved: (*a*) averaging the scores from several independent testing sessions, and (*b*) providing additional training materials for teachers.

The possibility of raising the reliability of quality scores by averaging the ratings for several independent samples is obvious. In view of the importance attached to evaluations of growth from

the beginning to the end of the year, it would be highly desirable to secure an average based on at least three test writings at each time of year. Additional steps could also be taken to improve the reliability of the scales. With more complete instructions on the emphasis to be given the various aspects of quality, teachers could probably develop more uniform standards in their judging. It seems likely that 2 or 3 hours of practice by a school faculty, using samples of pupil writing projected on a screen, would pay substantial dividends in increased reliability. This practice would be especially valuable for beginning teachers who have had little experience with handwriting scales. To reduce judge-to-judge differences in their overall leniency, publishers might consider supplying a set of practice specimens with each grade scale. These specimens, covering a suitably wide range of quality, could be reproduced on a set of cards, one sample to a card. On the reverse side of each card could be noted the average score assigned by the publisher's panel of judges. Such a practice deck could be quite helpful in bringing teachers closer together in their overall average ratings.

While this study has been concerned primarily with the issue of reliability, problems of validity clearly merit further investigation. The relationship between handwriting scale scores and quality measures derived from everyday writing samples should be considered. The content of the test sentences should be carefully analyzed to determine if current scales include those letter combinations, words, and numerals most sensitive to distortion of legibility. And of fundamental importance, of course, is the development of improved criterion measures of handwriting legibility.

REFERENCES

AYERS, L. P. A scale for measuring the quality of handwriting of school children. Bulletin 113, 1912, Division of Education, Russell Sage Foundation, New York. (Reprinted: Ayers, L. P. *Measuring Scale for Handwriting: Gettysburg Edition.* Cincinnati: Public School Publishing, 1915.)

FREEMAN, F. N. An analytical scale for judging handwriting. *Elem. school J.*, 1915, **15**, 432–441.

FREEMAN, F. N. *Guiding growth in handwriting.* Columbus, Ohio: Zaver-Bloser, 1959.

LINDQUIST, E. F. Estimation of variance components in reliability studies. In E. F. Lindquist (Ed.), *Design and analysis of experiments in psychology and education,* Boston: Houghton Mifflin, 1953. Ch. 16.

THORNDIKE, E. L. Handwriting. *Teachers Coll. Rec.*, 1910, **11**, 83–175.

(Received May 2, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 293–298

FAILURE IMAGERY IN THE FANTASY OF EIGHTH GRADERS AS A FUNCTION OF THREE CONDITIONS OF INDUCED AROUSAL

RICHARD C. ANDERSON
Harvard University [1]

An experiment was conducted in which 3 groups of 8th graders (total N = 143) completed a projective test under conditions of low-arousal, medium-arousal, and high-arousal respectively. Fear of Failure (f Failure), as well as n Achievement, significantly discriminated the 3 conditions of induced arousal during projective testing. The medium-arousal group rather than the high-arousal group yielded the highest mean f Failure. It was argued that negative results in past research on f Failure may have resulted from anxiety-provoked repression of failure imagery under high-arousal conditions.

The original research of the McClelland group (McClelland, Atkinson, Clark, & Lowell, 1953, 218–274) yielded evidence suggesting that there might be two recognizable aspects of achievement motivation, fear of failure (f Failure) as well as hope of success (n Achievement). Subsequent work has sustained this belief (Clark, Teevan, & Ricciuti, 1958). Nevertheless, in two investigations using a full-blown measure of f Failure (Moulton, 1958; Raphelson & Moulton, 1958) the f Failure means of achievement-oriented groups were not significantly higher than the f Failure means of relaxed groups. The research described in this paper was designed to see whether a projective measure of f Failure would discriminate conditions of induced arousal and to explore the relationships of f Failure to self-descriptive measures of anxiety.

PROCEDURE

The research was conducted with a sample of 143 eighth graders attending suburban junior high schools in the Boston area. The total sample was divided into three experimental groups. The intelligence test means of the three groups were not significantly different nor were the means on an index of socioeconomic status. The data of males and females were pooled since a pilot study yielded no significant sex differences.

The experiment was designed in such a way that projective protocols were elicited from three groups of subjects under three different conditions of arousal. All subjects performed on the first form of a digit-letter task under what might be called "neutral conditions." Following performance on the first form of the digit-letter task all groups were told that they had just done very poorly on a test of intelligence which was related to college success, driving skill, and being a good secretary. At this point, the second form of the digit-letter task was administered. One of the three experimental groups took the projective test just prior to the first form of the digit-letter task, under what can be called "low-arousal conditions." The test was introduced without special fanfare. Instructions contained no reference to "test." A second experimental group received the projective test under "medium-arousal conditions." The projective test was given prior to performance on the first form of the digit-letter task and was preceded by an "explanation" of the research in which the experimenter stated that he was "trying to find out how to make school more interesting and valuable" for bright, successful students and those who are "kind of stupid." The projective test was introduced as a "test of imagination" and the instructions contained several references to test. The "high arousal" group took the projective test following the second form of the digit-letter task. Once again the experi-

[1] The author is now with the East Brunswick Public Schools, New Jersey.

menter indicated that the subjects had done poorly on a test of important abilities. He then stated that he would give them the opportunity to redeem themselves on a test of imagination.

Four pictures, flashed on a screen in front of the room, were used to elicit projective protocols. The pictures can be briefly described as follows:

1. A Negro boy in an integrated class is raising his hand.

2. A boy leaning against a light pole is talking with a girl standing by a car.

3. In the foreground, a man and a boy kneel on the floor together, apparently doing homework, working on a puzzle, or playing a game. Two women and a baby are on a sofa in the rear.

4. A group of teenage boys and girls are sitting in a semicircle. They appear to be listening to what a boy in the foreground is saying.

Following usual procedure (Atkinson, 1958, p. 837), the four standard questions were printed on projective answer sheets to structure the subjects thinking slightly.

The scoring conventions recently revised by the McClelland group (McClelland, Atkinson, Clark, & Lowell, 1958) were used in determining n Achievement. These conventions were supplemented with the category Explicit Evaluation which Ricciuti and Clark (1957, p. 23) found to be a particularly good predictor of the presence or absence of achievement motivation. A revised and extended version of Moulton's schema was employed in scoring fear of failure. The chief revision was the adding of the "negative" categories (I−, Ga−, Bp, Bw, G−), which appeared in stories otherwise manifesting positive achievement motivation, to the f Failure score. Table 1 contains a brief summary of the f Failure scoring conventions used in the present investigation. For illustration, f Failure scoring has been spelled out in detail for the following story of an eighth grade boy:

> A teenage boy is called on in school and everyone is awaiting the outcome of his answers—the boy wasn't paying attention so everybody looks at him—they think the boy is stupid! The boy wants the answers from one of his classmates—the boy will get bawled out by the teacher and may be kicked out of class.

TABLE 1

SUMMARY OF f FAILURE SCORING SYSTEM

Name of category	Scoring weight	Brief definition
Failure imagery (FI)	2	Failure imagery is scored when there is (*a*) any indication of poor performance in customary achievement situation, (*b*) any doubt of ability to reach achievement goal, (*c*) any anticipation of failure, (*d*) any indication of discouragement or negative affect at the possibility of achievement evaluation, (*e*) an expressed preference to leave an achievement situation or situation in which achievement evaluation might take place, (*f*) daydreaming in an achievement situation unless the subject of the daydream is of successful achievement.
Doubtful failure imagery (DFI)	1	
Need to escape achievement situation (Ne)	1	Need to escape is scored when (*a*) a desire to leave an achievement situation is expressed or (*b*) a preference is expressed for another type of activity in an achievement situation, or (*c*) a desire for someone else to help or takeover in an achievement situation is expressed.

TABLE 1—Continued

Name of category	Scoring weight	Brief definition
Instrumental failure [a] (I—)	1	Instrumental failure is scored whenever some physical or mental act designed to reach a positively cathected achievement goal fails.
Instrumental escape (Ie)	1	Instrumental escape is scored whenever a character acts to avoid or escape an achievement situation.
Anticipation of failure [a] (Ga—)	1	Anticipation of failure is scored whenever a character doubts he will succeed or predicts failure.
Block in world to achievement success [a] (W—)	1	Scored whenever a physical object, turn of events, or other person stands in the way of attaining an achievement goal.
Block in person to achievement success [a] (P—)	1	Scored whenever some personal characteristic, such as negligence or laziness, stands in way of attaining achievement goal.
Block in world to escape from achievement situation (We)	1	Scored whenever some physical object, turn of events, or other person stands in way of escaping from achievement situation.
Block in person to escape from achievement situation (Pe)	1	Scored whenever some personal characteristic stands in way of escaping achievement situation.
Negative affect at failure [a] (G—)	1	Negative affect is scored when (*a*) someone experiences a negative emotional state as the result of poor performances or low evaluation. (*b*) the objective consequences of failure fall into the lower 10% of possible outcomes.
Hostile press (HosP)	1	(*a*) Hostile press is scored whenever there is overt criticism of one character by another, including predictions of failure, (*b*) ridicule or negative affect displayed toward character who fails.
Negative evaluation (E—)	1	Negative evaluation is scored whenever a character evaluates his own performance as inadequate.
Failure thema (FTh)	1	Failure thema is scored whenever a story meets the definition of failure imagery, contains no reference to positive striving for achievement success, and contains no major subplots manifesting other types of motivation.

[a] Categories which are scored, when present, in stories showing achievement imagery (AI) but not failure imagery (FI). Otherwise none of the categories is scored unless FI is first scored.

Failure imagery—There are several obvious grounds for scoring fear imagery. Having scored fear imagery it is possible to score other failure categories if they are present. Scored 2.

Need to escape achievement situation—"the boy wants the answers from one of his classmates." Scored 1.

Block in person to achievement success—"the boy wasn't paying attention." Scored 1.

Hostile press—"everybody looks at him . . . they think the boy is stupid . . . the boy will get balled out by the teacher." Scored 1.

Negative affect at failure—By analogy to the conventions developed in connection with n Achievement (McClelland et al., 1958, p. 200), if the objective state of affairs accompanying failure falls into the lower 10% of possible outcomes for that type of activity G— is to be scored. Applying this convention to the present story, it does seem that the consequences of failing to know an answer were especially harsh. Scored 1.

Failure thema—The story concerned failure and only failure, thus FTh must be scored. Scored 1.

Summing the scores assigned for individual categories, the subject received a f Failure score of 7 for this story. His total score consisted of the sum of his scores on all stories.

Two raters, with previous experience scoring n Achievement, independently scored the stories collected in the present investigation when they had reached a criterion of agreement (rho = .90) on sets of practice stories. All scoring was done without knowledge of which experimental group the subject was from, what score the subject had obtained on other stories, or any other information about the subject. One rater scored all of the stories written to the first and third pictures. The other scored the stories written to the second and fourth pictures. For reliability purposes, both raters scored all the stories of 40 randomly chosen subjects. Rank-order correlations of .76 and .89 were obtained between n Achievement scores and f Failure scores of the two raters, respectively. Stories written to one picture were discarded when it was discovered that only 3 of 143 protocols contained either achievement or failure imagery. So each subject's total n Achievement or f Failure score represented the sum of his scores on three stories. The distributions of raw n Achievement and raw f Failure scores were both positively skewed. The transformation $\sqrt{x+3}$ was applied to n Achievement scores while the transformation $\sqrt{x+1}$ was applied to f Failure scores. These transformations were applied before testing the significance of differences among means, or computing correlations with other variables.

All subjects completed a battery of six self-descriptive measures which included the Manifest Anxiety scale (Taylor, 1953) and a test anxiety scale.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF PROJECTIVE MEASURES FOR THREE EXPERIMENTAL GROUPS

Projective measure	Low-arousal group ($N = 47$)	Medium-arousal group ($N = 51$)	High-arousal group ($N = 45$)
n Achievement			
M	1.75	2.08	2.26
SD	.89	.94	.93
f Failure			
M	1.58	2.05	1.78
SD	.64	.79	.70

Note.—Statistics in this table are based on transformed data. Transformations of $\sqrt{x+3}$ and $\sqrt{x+1}$ were applied to n Achievement and f Failure, respectively.

RESULTS AND DISCUSSION

The n Achievement and f Failure means and standard deviations of the three experimental groups appear in Table 2. Bartlett's test indicated that the variances of the three groups were not significantly different on either of the projective measures. Analysis of variance (see Tables 3 and 4) showed the n Achievement means and the f

TABLE 3

ANALYSIS OF VARIANCE OF N ACHIEVEMENT MEANS OF THREE EXPERIMENTAL GROUPS

Source	*SS*	*df*	*MS*
Between	6.41	2	3.21
Within	120.02	140	.86
Total	126.43	142	

$$F = \frac{3.21}{.86} = 3.74^{*}$$

$^{*} p < .05.$

Failure means of the three groups significantly different at the .05 and .01 levels, respectively. The total-sample correlation between n Achievement and f Failure was .19.

In the case of n Achievement, the means of both the high-arousal and medium-arousal groups were significantly higher than the mean of the low-arousal group. The high-medium difference was not significant.

The f Failure mean of the medium-arousal group, but not the high-arousal group, was significantly higher ($p < .001$) than that of the low-arousal group. The f Failure mean of the medium-arousal group was higher than the f Failure mean of the high-arousal group, but when a two-tailed test was applied the difference fell short of customary levels of significance ($.10 > p > .05$). Why did the medium-arousal group display the most failure imagery? The high-arousal group had just "failed" on an "intelligence test" which the group was told was related to getting into college, being a good secretary, and driving a car. Yet the medium-arousal group, in which achievement was made salient in a much less forceful and personal way showed a higher f Failure mean than the high-arousal group. Conceivably, conditions of high-arousal provoke anxiety about failure and result in a repression of failure imagery among some subjects. In a less personal achievement-relevant situation—such as the medium-arousal condition—anxiety is not aroused to such a degree and defenses are relatively inoperative, so considerable failure imagery does appear. In a low-arousal situation where there are few achievement-related cues little failure imagery appears.

In short, a curvilinear relationship is postulated between degree of arousal in the projective testing situation and amount of failure imagery appearing in projective protocols. This hypothesis may account for the fact that the medium-arousal group showed the highest mean f Failure in the present experiment and may also account for the negative results of previous research (Moulton, 1958; Raphelson & Moulton, 1958), if it is assumed that the more arousing of the experimental conditions in these studies provoked anxiety about failure.

It is argued that the high-arousal group displayed less f Failure than the medium-arousal group, and only slightly more than the low-arousal group, because of a defensive reaction to an anxiety-provoking situation. The pattern of relations f Failure from the various arousal groups showed with the two self-descriptive anxiety scales is congruent with this interpretation. High-arousal f Failure correlated $-.11$ with the Manifest Anxiety scale and $-.22$ with the Test Anxiety Scale. In contrast, low-arousal f Failure correlated .18 and .19 with the Manifest Anxiety scale and the Test Anxiety Scale, respectively. The parallel coefficients for the medium-arousal group were .03 and .10. None of these correlations alone was significantly different from zero, nor were differences between any of the correlations significant. However, the direction of the coefficients is congruent with the hypothesis that anxious persons in the high-arousal group repressed failure imagery.

TABLE 4

ANALYSIS OF VARIANCE OF THE F FAILURE MEANS OF THREE EXPERIMENTAL GROUPS

Source	*SS*	*df*	*MS*
Between	5.63	2	2.81
Within	73.31	140	.52
Total	78.94	142	

$$F = \frac{2.81}{.52} = 5.40^{*}$$

$^{*} p < .01$.

These results are consistent with those reported by Raphelson and Moulton (1958). In their first experiment, a red light flashed when subjects made errors on a pursuit rotor, which they frequently did. Fear of Failure, scored from a projective test administered immediately after performance on the pursuit rotor, correlated .25 with the test anxiety questionnaire. In the second experiment the flashing light was omitted. Raphelson and Moulton argued that the lack of information in the second experiment increased uncertainty about performance, and that "this perceived uncertainty" increased "the amount of aroused anxiety in all Ss." In the second experiment, f Failure correlated −.38 with the test anxiety questionnaire. The difference between the correlations obtained in the two experiments was significant at the .05 level. Raphelson and Moulton interpreted the negative correlation obtained in the second experiment as meaning that highly anxious subjects were defensive or "left the field psychologically." The similarity in findings of the present investigation and the Raphelson-Moulton studies, despite different experimental conditions, lends credibility to the results and their interpretation.

REFERENCES

ATKINSON, J. W. (Ed.). *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

CLARK, R. A., TEEVAN, R., & RICCIUTI, H. N. Hope of success and fear of failure as aspects of need for achievement. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton: Van Nostrand, 1958.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A. & LOWELL, E. L. A scoring manual for the achievement motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

MOULTON, R. W. Notes for a projective measure of fear of failure. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton: Van Nostrand, 1958.

RAPHELSON A. C., & MOULTON, R. W. The relation between imaginative and direct verbal measures of test anxiety under two conditions of uncertainty. *J. Pers.*, 1958, **26**, 556–567.

RICCIUTI, H. N., & CLARK, R. A. *A comparison of need-achievement stories written by experimentally "relaxed" and "achievement-oriented" subjects: Effects obtained with new pictures and revised scoring categories*. Princeton: Educational Testing Service, 1957.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.

(Received May 7, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 299–302

TEST ANXIETY, EXPERIMENTAL INSTRUCTIONS, AND THE WECHSLER ADULT INTELLIGENCE SCALE [1]

IRWIN G. SARASON AND JAMES MINARD [2]
University of Washington

4 subscales of the WAIS were administered to 96 college students in an experiment in which there were 3 variables: (a) Ss' Test Anxiety Scale scores, (b) sex of S, and (c) preliminary conditions (achievement-orienting and neutral instructions). On the Vocabulary, Block Design, and Comprehension subscales low test anxious Ss were superior to high test anxious Ss. For Comprehension, this effect was largely attributable to the effects of instructions. High and low test anxious Ss did not differ on Digit Symbol, but a Test Anxiety × Sex × Instructions interaction was obtained for this subscale. The results were interpreted as indicating the relevance to performance of the 3 variables studied as well as the importance of the task variable in research and anxiety.

The present paper reports evidence relevent to two questions: Do subjects differing in test anxiety differ in their performance on an individually administered standardized intelligence test, a short form of the Wechsler Adult Intelligence Scale (WAIS)? Do experimental orienting instructions administered prior to performance differentially affect high and low test anxious subjects?

The experiment performed was conceived in relation to certain empirical findings already reported (Sarason, 1960). Available evidence has demonstrated a significant negative relationship between scores on the Test Anxiety Scale (TAS) and group administered indices of intelligence (Sarason, 1957, 1959, 1961). There are also indications reported of differential influences of stressful and nonstressful instructions on the performance of subjects differing in anxiety (Lucas, 1952; Nicholson, 1958; Sarason, 1956). These latter findings have come, however, primarily from verbal learning experiments. It seemed of value to determine the relationship of TAS scores to the individually administered WAIS and, also, to determine the extent of the interaction between TAS scores and instructions for that test.

METHOD

Subjects

The subjects were 96 college students enrolled in introductory psychology classes. They had been selected from a larger group of students, which had been administered the true-false TAS (Sarason, 1959, 1960; Sarason & Palola, 1960). Half of the 96 subjects had TAS scores of 8 and above, and half had scores of 0 or 1. These cutoff points corresponded approximately to the upper and lower 15% of the TAS distribution. The high and low TAS groups were selected so that each was composed equally of males and females.

Procedure

Each subject was individually administered four subtests of the WAIS (Wechsler, 1955). Two Verbal scale subtests, Vocabulary and Comprehension, and two Performance scale subtests, Block Design and Digit Symbol, were the tasks employed. Except for the experimental preliminary instructions, the testing procedures outlined in the WAIS manual were followed.

The order of presentation of the four subtests was handled in the following manner.

[1] This study was supported by a grant from the Agnes H. Anderson Research Fund.

[2] Now at the University of Colorado.

Half of the subjects were administered the two Verbal tests first, while the remaining subjects were administered the Performance tests first. Order of administration within the two Verbal and within the two Performance tests was randomized. In order to determine whether or not order of administration of Verbal and Performance subtests influenced the subject's WAIS scores, *F* tests were performed. The results showed that there was no significant order effect.

Half the subjects received the following achievement-orienting instructions at the beginning of the experiment:

> In this experiment I want to make an evaluation of your intellectual level. Although most students in college are above average intelligence there are some who aren't any better than average in ability. By using the scales of the Wechsler Adult Intelligence Test, we will be able to pinpoint those students who are below the level of intelligence needed for successful work in college.

The other subjects received the following more neutral instructions:

> I'd like you to help me in this way. These are some tasks I'm getting practice in administering. They start with easy items and work up to more difficult ones.

RESULTS

Three factor analyses of variance (12 subjects per cell) were performed using subtest raw scores and the scaled score sums as the measures. Table 1 presents the means for the scaled score sums and for the subtest raw scores. When the sum of the scaled scores was used as the measure, only one *F* was found to be statistically significant. Low TAS subjects obtained higher scores than did high TAS subjects ($p < .05$). Analysis of various of the subtest raw scores for Vocabulary yielded the same results ($p < .05$). In both these analyses, the superiority of low to high TAS subjects was as great under the neutral as under the achievement-orienting condition.

For Comprehension, the low TAS group performed at a higher level than did the high TAS group ($p < .05$). However, this result was due largely to differences under the achievement-orienting condition. Under this condition, low TAS subjects were superior to high TAS subjects ($p < .02$), but under the neutral condition there was no significant difference.

The Digit Symbol analysis did not reveal any overall differences due to TAS or instructions. The one significant *F* involved the interaction of TAS scores, instructions, and sex ($p < .05$). This result was due to differences in response to experimental instructions of male and

TABLE 1

MEANS FOR SUM OF SCALED SCORES AND FOR SUBTEST RAW SCORES (12 *Ss* per group)

Group [a]	Scaled scores sum	Comprehension	Vocabulary	Digit Symbol	Block Design
H-M-(A-O)	53.83	23.92	62.42	65.00	39.42
H-F-(A-O)	52.50	23.25	61.92	65.42	36.33
H-M-N	52.67	24.00	63.25	55.75	38.42
H-F-N	54.42	24.33	62.00	68.67	26.17
L-M-(A-O)	56.92	26.00	70.00	62.58	42.83
L-F-(A-O)	56.08	25.42	64.00	65.75	37.92
L-M-N	57.25	24.42	65.50	64.50	44.17
L-F-N	54.75	24.42	67.08	63.50	37.08

[a] First designation characterizing each group refers to high (H) or low (L) TAS scores; second designation refers to male (M) or female (F) subjects; and third designation indicates either achievement-orienting condition (A-O) or neutral condition (N).

female high TAS subjects. Female high TAS subjects performed at a higher level under neutral than under achievement-orienting conditions, whereas male high TAS subjects performed at a higher level under the achievement-orienting than under the neutral condition.

There were two significant *F*s in the analysis of Block Design scores. One was for sex ($p < .01$), with males superior to females; the other was for TAS ($p < .05$), with low superior to high TAS subjects. This difference between high and low TAS subjects tended to be stronger for males than for females.

DISCUSSION

With respect to overall performance level, it seems clear that low test anxious subjects were generally superior on this short form of the WAIS to high test anxious subjects. This result confirms previous findings using group administered measures of intelligence (Sarason, 1957, 1959). How may these results be interpreted? Two alternatives come to mind. Perhaps TAS is simply a measure of intelligence. On the other hand, it might be considered as a measure of a nonintellective, personality variable which, with intelligence, influences performance. Obviously the latter hypothesis is, psychologically, the more interesting and potentially more fruitful conception. Supporting this hypothesis is evidence from learning experiments which has indicated that under certain instructional conditions low anxious groups are superior to high anxious groups, and under other conditions there are no significant differences (Sarason, 1956, 1958). This suggests that ability, per se, may not be the basis for observed differences between high and low anxious groups; and that these groups differ in their response to the variable of instructions.

One way of viewing TAS scores is to assume that they reflect the tendency of the individual to show deterioration in performance under personally threatening or stressful conditions which are associated with performance. Or, as Atkinson (1960) has suggested, anxiety scores may reflect varying strengths of the need to avoid failure. This interpretation would lead one to view experimental instructions and conditions as important factors in the arousal of the need. If this type of assumption is correct it might, then, be hypothesized that (*a*) under such conditions, high TAS subjects would be inferior in performance to low TAS subjects, but that (*b*) under more neutral conditions, there would be no difference attributable to TAS score differences.

Of the four tasks used in the present experiment, this hypothesis was supported in the case of Comprehension, but not in the cases of Vocabulary and Block Design. Support for the hypothesis of a relationship between anxiety and experimental instructions was obtained for Digit Symbol scores, but within the context of an interaction more complex in nature than would be encompassed by the conception suggested above. For Digit Symbol, the TAS-instructions interaction was complicated by the variable of sex of the subject. For females, as would be predicted, high TAS subjects performed at a higher level under achievement orientation than under neutral conditions. However, the reverse situation, obtained for high TAS males, had not been expected.

The lack of significant simple relationships for anxiety and instructions for Digit Symbol and the obtained triple interaction is interesting in the light of previous research in which these variables were found to interact significantly with a third variable. Sarason and Palola (1960), using a digit sym-

bol task, found that TAS and instructions interacted significantly with task difficulty. The present results, together with evidence already reported, suggest that the relationship between anxiety and experimental instructions may be importantly influenced by the nature of the task on which the subjects perform, the difficulty of that task, and, at least one subject characteristic, sex. The nature of the task seems to be particularly important, since the degree of effectiveness of experimental instructions which describe the task as a measure of intelligence probably depends on the extent to which the task has face validity for the subject as a measure of intelligence. Both instructions and the task can be viewed as inducing varying degrees of motivation. Research incorporating these, and other potentially relevant factors, such as characteristics of experimenters and the particular anxiety instrument employed, should prove valuable in clarifying the relationship of anxiety to performance. This sort of research would also be relevant to the clinical testing situation since it would shed light on the complex question of the optimum method of presenting or introducing psychometric instruments to individuals.

REFERENCES

ATKINSON, J. W. Personality dynamics. *Annu. Rev. Psychol.*, 1960, **11**, 255–290.

LUCAS, J. D. The interactive effects of anxiety, failure, and intraserial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.

NICHOLSON, W. M. The influence of anxiety upon learning. *J. Pers.*, 1958, **26**, 303–319.

SARASON, I. G. Effect of anxiety, motivational instructions, and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–260.

SARASON, I. G. Test anxiety, general anxiety, and intellectual performance. *J. consult. Psychol.*, 1957, **21**, 485–490.

SARASON, I. G. Effects on verbal learning of anxiety, reassurance, and meaningfulness of material. *J. exp. Psychol.*, 1958, **56**, 472–477.

SARASON, I. G. Intellectual and personality correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1959, **59**, 272–275.

SARASON, I. G. Empirical findings and theoretical problems in the use of anxiety scales. *Psychol. Bull.*, 1960, **57**, 403–415.

SARASON, I. G. Test anxiety and the intellectual performance of college students. *J. educ. Psychol.*, 1961, **52**, 201–206.

SARASON, I. G., & PALOLA, E. G. The relationship of test and general anxiety, difficulty of task, and experimental instructions to performance. *J. exp. Psychol.*, 1960, **59**, 185–191.

WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation, 1955.

(Received May 23, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 303–309

INFLUENCES ON THE STUDENT'S MOTIVATION TO SEEK ADVANCED TRAINING: ANOTHER LOOK [1]

ALEXANDER W. ASTIN
National Merit Scholarship Corporation, Evanston, Illinois

A recent study (Thistlethwaite, 1962), in which certain college "presses" were interpreted as influencing the student's motivation for advanced training, was re-examined. The method used for controlling the student's initial educational level of aspiration—the analysis of covariance—was thought to be inappropriate, primarily because of anomalies in the correlation between initial and final levels of aspiration. In addition, the estimate of initial level of aspiration, which had been obtained retrospectively from the Ss, was judged to be too undependable for use as a measure of initial level of aspiration. Reanalyses of some of the data failed to support the author's original conclusions about the "effects" of the press on student motivation. Certain methodological problems in studying the influence of college environments on the student were discussed.

In a recent paper Thistlethwaite (1962) reported an extensive series of analyses which appear to demonstrate the influence of faculty and student "presses" on the college student's educational plans. He concluded from these analyses that the male student's level of aspiration (LOA, highest degree planned) is increased during the first 2 years of college by high faculty presses for "independence," "supportiveness," and "affiliation"; low faculty press for "vocationalism"; high student press for "openness to faculty influence"; and participation in honors programs. The principal justification for these interpretations was that the analyses had ruled out certain "rival hypotheses" (i.e., rival to the "environmental effect" hypothesis), namely, the student's sex, major field, and level of aspiration at the time he enters college.[2] However, a re-examination of the evidence suggests that these conclusions are vitiated by inaccuracies in the student reports of initial LOA and by the method of analysis which was used.

Appropriateness of the Method

The principal statistical method used by Thistlethwaite was the analysis of covariance. Posttest LAOs (after 2 years of college) for various "press" groups were adjusted in terms of pretest LOAs (obtained retrospectively). Of the several assumptions required by the analysis of covariance, the nature of the regression between the control

[1] This paper is a part of the research program of the National Merit Scholarship Corporation and was financed by a grant from the National Science Foundation. The author is indebted to J. L. Holland for his critical reading of the manuscript.

[2] It should be noted that these rival hypotheses, which Thistlethwaite attributes to the present writer (Astin, 1961) and others (Holland, 1957; Stuit, Helmstadter, & Frederiksen, 1956), were not proposed with respect to the "student's stated level of aspiration after two years of college." On the contrary, these hypotheses were proposed as alternatives to an environmental effects interpretation of a quite different dependent variable: the proportion of a college's graduates who eventually obtain the PhD.

and dependent variables is of special importance. Figure 1 shows the distribution of Posttest LOAs for each level of Pretest LOA using Thistlethwaite's male subjects. (The lowest level, "less than bachelor's degree," has been omitted because only one subject gave this as his Pretest LOA.) An inspection of Figure 1 indicates that the regression is heteroscedastic. This observation is supported by Bartlett's test, which is significant ($p < .01$). In fact, four of the six possible pairs of column variances are heterogeneous at the .01 level. In addition to heteroscedasticity, several of the arrays in Figure 1 are markedly skewed, and two of the arrays appear to have bimodal distributions.

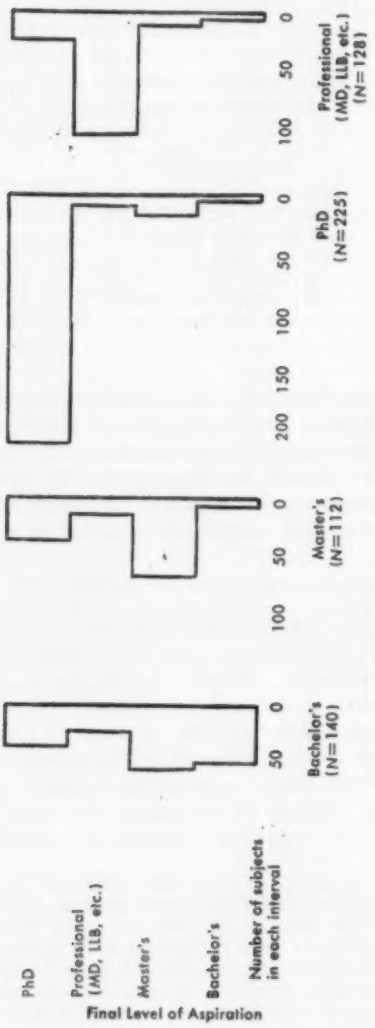

FIG. 1. Final LOA after 2 years of college as a function of retrospective report of initial LOA at time of entrance to college. (Male subjects from Thistlethwaite, 1962).

In view of these deviations from the conditions necessary for employing the Pearson r, it would seem that controlling initial LOA by means of the analysis of covariance is inappropriate. In addition, since initial LOA was obtained from the subject concurrently with a report of the press, the possibility exists that the independent variable (press) has affected the control variable, initial LOA, (a discussion of how this might occur follows later). Such an effect would also invalidate the analysis of covariance.

If one were to overlook the anomalies in Figure 1, the use of covariance analysis in this situation can still be questioned on the grounds that similar "adjusted posttest" LOAs are in some cases based on qualitatively different events. For instance, the student who changes from PhD to the MD degree receives approximately the same adjusted posttest LOA as the student who changes from the MA to the BA degree. A related question is whether the MD or LLB actually represent "lower" levels of aspiration than the PhD. Does the student who abandons his plans for the MD and pursues instead the PhD in biology really "raise" his level of aspiration? The MD is no lower in terms of the amount of formal training required, and a good case can probably be made for the point of view that the

MD has more prestige associated with it than almost any other degree. In terms of the academic degree hierarchy, the MD and LLB rank with the baccalaureate degrees; but to follow this ordering would be to treat any professional degree as a lower LOA than even the master's degree. Perhaps the most defensible ordering, in terms of status and length and difficulty of training required, would be to combine the PhD and professional degrees into a single ("highest") category.

ACCURACY OF THE RETROSPECTIVE REPORT

All analyses of the effects of different presses were based on retrospective reports by the subjects of their LOAs at the time they entered college. While it was acknowledged that the true initial LOA "possibly . . . would not correlate perfectly with retrospective pretests" (Thistlethwaite, 1962, p. 54), certain other analyses recently completed with another sample suggest that such retrospective reporting is highly inaccurate. These analyses, based on a large sample of 1957 Merit Finalists and Commended Students, are shown in Table 1. These students were asked their LOA shortly after they entered college in Fall 1957, and again as they were completing college in 1961. In the 1961 survey they were also asked to recall what their LOA had been in 1957. From Table 1 it is clear that errors in the retrospective report are common. There is a systematic tendency for both males and females to underestimate their initial LOA, although overestimation also occurs with some frequency. Errors in recall vary with each actual level in 1957, the most frequent occurring at the master's level, where only 33% of the students who initially chose the master's level in 1957 "remembered" in 1961. Accuracy of recall among students who initially chose the bachelor's or doctor's degree in 1957 was 79%.

A further analysis of the subjects in Table 1 showed that among the 1,605 males who *changed* LOA between 1957 and 1961, only 45% recalled their initial LOA correctly. The accuracy of recall for the 2,336 male nonchangers was 79%. Accuracy of recall was 57% for the 874 female changers and 70% for the 1,029 female nonchangers. A possible explanation for this differential accuracy between changers and nonchangers is that when the subject is unsure about what his initial LOA really

TABLE 1

ACCURACY OF THE RETROSPECTIVE REPORT: ACTUAL LOA IN 1957 COMPARED WITH RETROSPECTIVE REPORT 4 YEARS LATER

Actual LOA in 1957	Retrospective report of 1957 LOA									
	Less than bachelor's		Bachelor's		Master's		Doctor's (PhD, MD, etc.)		Totals	
	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
Doctor's (PhD., MD., etc.)	1	0	222	47	168	46	1633	224	2024	317
Master's	1	0	493	304	362	235	294	111	1150	650
Bachelor's	0	5	606	775	115	127	82	36	803	943
Less than bachelor's	1	5	11	36	2	3	6	2	20	46
Totals	3	10	1332	1162	647	411	2015	373	3997	1956
Percentage of correct recall	33.3	50.0	45.5	66.7	56.0	57.2	81.0	60.1	65.1	63.3

was, he tends to give his current LOA. Whenever this happens the nonchanger would be "right" and the changer "wrong."

Although it is not possible to check on the accuracy of recall of LOA for all of the subjects used by Thistlethwaite, some of his sample does overlap with another sample of Merit Finalists (Holland, 1960) on which LOA was obtained immediately prior to their entering college in 1958. Since Thistlethwaite's positive results with environmental effects were obtained with male subjects,[8] the following analyses are confined to the 150 male subjects who overlapped in the two samples. Table 2 shows the accuracy of the retrospective report of LOA obtained from these subjects by Thistlethwaite. The pattern of errors in Table 2 is very similar to that obtained with males in the much larger sample (Table 1), even though the interval of time was only 2 (rather than 4) years. The slightly greater rate of error for Thistlethwaite's males may be due in part to the use of the additional level (professional degrees). Once again, the changers ($N=57$) showed a much lower accuracy of recall (26%) than the nonchangers (81%).

It seems clear from these results that retrospective reports of LOA, even after only 2 years, are undependable. The rate of accuracy for the subjects who actually *do* change LOA is essentially chance. Assuming that *no* subjects changed LOA between 1958 and 1960 results in fewer errors (62% accuracy) than were involved in the retrospective reports (60% accuracy).

In short, the evidence suggests that retrospective reports of initial LOA are not sufficiently accurate to use as substitutes for true initial LOA. Accordingly, this and other studies (e.g., Thistlethwaite, 1960) which have utilized such retrospective reports require replication with the use of true initial and final LOAs before one can conclude that the college press affects LOA.

REANALYSIS OF DATA

Since there were 150 male subjects for whom initial statements of LOA in 1958 were available, an attempt was made

[8] Thistlethwaite does mention one significant environmental effect on LOA among women: the Faculty Press for Humanism was negatively related to increased LOA. This finding was exactly the opposite of the effect of the Humanism press reported in a previous study (Thistlethwaite, 1960). Both of these studies, however, used retrospective reports of initial LOA.

TABLE 2

ACCURACY OF THE RETROSPECTIVE REPORT FOR 150 MALE SUBJECTS
(Subjects from Thistlethwaite, 1962)

Actual LOA in 1958	Retrospective report of 1958 LOA				
	Bachelor's	Master's	Professional (MD, LLB, etc.)	PhD	Totals
PhD	11	9	8	51	79
Professional (MD, LLB, etc.)	2	0	17	5	24
Master's	10	10	0	6	26
Bachelor's	12	3	2	4	21
Totals	35	22	27	66	150
Percentage of correct recall	34.3	45.5	63.0	77.3	60.0

to replicate Thistlethwaite's positive findings regarding the effects of press. The first problem was to determine whether any of the relationships between press and final LOA held up when initial (true) LOA was held constant. For this purpose three separate analyses were carried out. In the first analysis all subjects whose initial LOA was at the bachelor's level were divided into two groups: those who raised their LOAs between 1958 and 1960 and those who did not. For the second analysis, the subjects with initial LOAs at the master's level were similarly divided into those who did not change and those who increased their LOAs. (In these first two analyses four subjects who lowered their LOAs, and two others who gave unclassifiable LOAs in 1960, were omitted. In addition, one subject, not shown in Table 2 because his retrospective LOA was "unclassifiable," was added.) In the third analysis, all subjects whose initial LOAs in 1958 were at the doctorate level (PhD, MD, etc.) were divided into two groups: those who maintained this LOA in 1960 and those who reported a lower LOA (i.e., master's or bachelor's) in 1960. One subject whose 1960 LOA was unclassifiable was omitted from this analysis.) The three groups of changers were then compared with their respective nonchangers on each of the presses by means of t ratios. Table 3 shows the results of these analyses. None of the 18 one-tailed t tests in Table 3 reaches significance. Exactly half of the mean differences are in the predicted direction, and the 2 largest t ratios occur with mean differences which are actually in opposite directions from what would be predicted from Thistlethwaite's conclusions.

TABLE 3

REANALYSES OF THE EFFECTS OF PRESS
(145 male subjects from Thistlethwaite, 1962)

Press variable	Initial (true) LOA in 1958: Bachelor's: Mean for nonchangers (N=6)	Bachelor's: Mean for subjects with higher LOA in 1960 (N=14)	Bachelor's: t [a]	Master's: Mean for nonchangers (N=12)	Master's: Mean for subjects with higher LOA in 1960 (N=11)	Master's: t [b]	Doctor's (PhD, MD, etc.): Mean for nonchangers (N=89)	Doctor's (PhD, MD, etc.): Mean for subjects with lower LOA in 1960 (N=13)	Doctor's (PhD, MD, etc.): t [c]
Faculty press:									
Independence (+)[d]	7.00	6.50	.46	5.75	6.45	.64	6.70	6.69	.00
Supportiveness (+)	7.67	7.29	.35	6.33	6.91	.57	6.61	6.46	.22
Affiliation (+)	8.50	6.29	1.71	6.50	6.36	.13	6.53	7.08	.71
Vocationalism (−)	7.83	6.64	.83	5.50	6.64	.87	5.74	6.69	1.20
Student press:									
Openness to faculty influence (+)	6.33	4.36	1.71	6.00	5.91	.11	5.69	5.31	.59
Honors program participation (+)[e]	16.7	14.3	.14	25.0	36.4	.59	40.4	30.8	.70

[a] $t_{.05} = 1.73$, one-tailed test.
[b] $t_{.05} = 1.72$, one-tailed test.
[c] $t_{.05} = 1.66$, one-tailed test.
[d] Direction of relationship between press and tendency to raise LOA as reported by Thistlethwaite (1962).
[e] The percentage of honors program participants in each group is shown.

In view of these negative findings,[4] the question naturally occurs as to what factors were involved in the significant relationships reported by Thistlethwaite (1962) using retrospective LOA and the analysis of covariance. He found (p. 59) that covarying on initial (retrospective) LOA did not remove all of the variance associated with the subject's initial major field. As a means of further controlling this "input" variable, the subjects were then separated into science and nonscience majors. On this basis, the Faculty Press for Independence was still significantly related to adjusted mean level of aspiration within the nonscience group. (Results of the science, nonscience analysis for the five other presses shown in Table 3 were not reported, so it is assumed that the relationships disappeared with the addition of this third input variable.) However, certain other results (Thistlethwaite, 1962) suggest that the division of subjects into science and nonscience groups may not have been a sufficient control for initial major field. It happens that two of the specific major field groups included in the nonscience classification, philosophy and education, had, respectively, the highest and lowest adjusted mean LOAs of all the specific major field groups (pp. 58–59). In Thistlethwaite's (1962) Table 1 (p. 56), which shows the relationships between major fields and some of the presses, it can be seen that philosophy—the group with the highest adjusted mean LOA—also has one of the highest mean scores on Faculty Press for Independence. Similarly, education—the group with the lowest adjusted mean LOA—has the lowest mean of all the nonscience groups on Faculty Press for Independence. Perhaps the most likely explanation of the significant relationships involving this press is that there were still differences with respect to initial major field of study which had not been adequately controlled.

[4] These reanalyses were performed with only 24% of Thistlethwaite's male sample. Although there is no reason to believe that the smaller sample is a biased one, the possibility remains that the results would have been different if the entire sample had been used.

DISCUSSION

The present results suggest that the college press does not affect the student's level of aspiration in the manner previously supposed. On the other hand, these findings should not be interpreted as evidence that the college environment is unimportant. Indeed, on strong a priori grounds we should expect students to be affected differently by different college environments. But since we cannot assign students randomly and manipulate their college environments directly, we cannot conclude that differential environment influences have been identified until we have adequately controlled those input variables which are known to be related to the student performance in question.

The importance of such controls can hardly be overestimated. First of all, since the college press (Pace & Stern, 1958) is closely related to certain characteristics of the student body (Astin & Holland, 1961), the effects of these presses cannot be properly assessed without first controlling for student input. Secondly, the control of student input should eventually lead us closer to whatever "real" effects might exist. Recent research on the PhD "productivity" of colleges (Astin, 1962), for example, has shown that some of the colleges which were previously called "highly productive" (Knapp & Greenbaum, 1953) turn out to be among the most "underproductive" when certain aspects of their student inputs are taken into account.

A final methodological point regarding studies of this type should be mentioned. If we seek to learn what aspects of the college environment influence the student, we should try to keep our measures of the environment and our measures of student performance experimentally independent. Thistlethwaite (1962) speaks of "students exposed to some educational treatments" and of "teachers who exert press for independence" (p. 63), almost as if the press were an experimentally manipulated independent variable. The fact is, that the press for any student was defined solely in terms of the student's own verbal report. It is possible, for example, that students who regard their professors positively would be more inclined to attribute beneficial effects to these same professors than students who are prone to be critical of their professors. To be sure, such effects might occur, but they will be difficult to demonstrate in any convincing fashion as long as the student defines his own status simultaneously on both the independent and dependent variables.

REFERENCES

ASTIN, A. W. A re-examination of college productivity. *J. educ. Psychol.*, 1961, **52**, 173–178.

ASTIN, A. W. "Productivity" of undergraduate institutions. *Science*, 1962, **136**, 129–135.

ASTIN, A. W., & HOLLAND, J. L. The Environmental Assessment Technique: A way to measure college environments. *J. educ. Psychol.*, 1961, **52**, 308–316.

HOLLAND, J. L. Undergraduate origins of American scientists. *Science*, 1957, **126**, 433–437.

HOLLAND, J. L. The prediction of college grades from personality and aptitude variables. *J. educ. Psychol.*, 1960, **51**, 245–254.

KNAPP, R. H., & GREENBAUM, J. J. *The younger American scholar: His collegiate origins*. Chicago: Univer. Chicago Press, 1953.

PACE, C. R., & STERN, G. C. An approach to the measurement of psychological characteristics of college environments. *J. educ. Psychol.*, 1958, **49**, 269–277.

STUIT, D. B., HELMSTADTER, G. C., & FREDERIKSEN, N. *Survey of college evaluation methods and needs*. Princeton: Educational Testing Service, 1956.

THISTLETHWAITE, D. L. College press and changes in study plans of talented students. *J. educ. Psychol.*, 1960, **51**, 222–234.

THISTLETHWAITE, D. L. Fields of study and development of motivation to seek advanced training. *J. educ. Psychol.*, 1962, **53**, 53–64.

(Received May 23, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 310–315

RIVAL HYPOTHESES FOR EXPLAINING THE EFFECTS OF DIFFERENT LEARNING ENVIRONMENTS[1]

DONALD L. THISTLETHWAITE
Vanderbilt University

2 explanations advanced to show the artifactual nature of previously reported effects of different learning environments are critically examined. Longitudinal data on a sample of 2405 undergraduate men confirm and extend the previous findings and rule out interpretation of the findings in terms of retrospective errors of recall or in terms of alleged anomalies related to covariance analysis. Men who report that their teachers exert strong press for enthusiasm, humanism, affiliation, independence, achievement, and supportiveness, or who exert weak press for compliance, tend to raise their aspirations for advanced training more than men not reporting such press.

In his paper Astin (1962) has questioned whether my earlier report (Thistlethwaite, 1962) has provided evidence that college press influences the student's motivation to seek advanced training. Astin proposes two hypotheses as alternative explanations for my results: (*a*) the observed differences were spurious due to inaccuracies in the retrospective recall of degree aspirations at the beginning of college, (*b*) the observed differences were spurious due to an inappropriate application of covariance analysis.

I shall show that neither of these rival hypotheses is plausible because of weaknesses in the argument and analysis presented by Astin and because neither receives support from more recent results obtained by the author. With regard to the first of Astin's rival hypotheses I shall argue: (*a*) his data do not provide convincing evidence for the inaccurracy of retrospective recall; (*b*) he has failed to show how alleged errors of recall could account for my earlier results; and (*c*) the alleged errors of recall are irrelevant in any case, since when they are ruled out the previously reported effects are replicated.

With regard to the second rival hypothesis I shall argue: (*a*) Astin has failed to show how my use of covariance analysis could have produced spurious results; (*b*) no valid inferences concerning this hypothesis can be drawn from the results of Astin's "reanalysis" (in Table 3 of his paper) because of sampling considerations; and (*c*) the results of a new analysis, based upon a large sample of undergraduates, provide no support for it.

Alleged Errors of Recall

Anyone familiar with the mail surveys made of Merit Finalists by the National Merit Scholarship Corporation knows that the returns typically come back over a 6-month interval or more. Consequently precise localization of the time period to which student reports refer is impossible. However, such lo-

[1] The research briefly reported herein was supported through the Cooperative Research Program of the Office of Education, United States Department of Health, Education, and Welfare. In addition, the collection of longitudinal data was partially supported by National Science Foundation, Old Dominion Foundation, and Ford Foundation grants to the National Merit Scholarship Corporation.

calization is critical for the kind of comparisons Astin attempts to make. In describing the procedures used in polling the 1957 Finalists and commended students, Astin (1962) states:

> These students were asked their LOA [level of aspiration] shortly after they entered college in Fall 1957, and again as they were completing college in 1961. In the 1961 survey they were also asked to recall what their LOA had been in 1957 (p. 305).

No matter how adroit the request for retrospective recalls in 1961, it is unlikely that each student was being asked to recall what his dispositions were during the *same time interval* to which his report in 1957 referred. In short, if there was any ambiguity regarding the specific point in time students were being asked to recall, then discrepancies of the kind Astin points to do not necessarily show errors of recall. Since there are ample grounds for assuming such ambiguity existed, it follows that Astin's (1962) data summarized in Tables 1 and 2 of his paper do not provide convincing evidence of the unreliability of retrospective reports.[2]

IMPORT OF ALLEGED ERRORS OF RECALL

Even if we grant that Astin's data raise questions about the reliability of retrospective reports, it does not follow that such unreliability could explain the observed differences. On the contrary, unreliability of measurement should work *against* the demonstration of the hypothesized effects. The more unreliability in the data the harder it should be to demonstrate significant differences between press groups. If effects are obtained with unreliable measures, they should be even greater with more reliable measures.

To be sure, Astin argues that the "changers" showed more inaccurate recall than the "nonchangers." However, my comparisons were not among groups dichotomized in this manner. Furthermore, no attempt has been made by Astin to show how the alleged systematic biases in recall could produce the "effects" observed in my earlier study.

COVARIANCE ANALYSIS WITH ERRORS OF RECALL RULED OUT

Even though we cannot imagine *how* inaccurate recall could have produced spurious results in the earlier study, it is worth asking whether the "effects" of different learning environments would disappear if potential errors of recall were ruled out. Table 5 of my earlier report (Thistlethwaite, 1962) describes the observed effects when the covariate (initial LOA) is based upon data obtained by the method of retrospective recall. A new analysis in which the covariate (initial LOA) is based on "actual" measures obtained at the beginning of college is reported below. If the latter analysis fails to confirm the former, then we shall have evidence that errors of recall may have somehow spuriously produced the effects. On the other hand, if the latter analysis confirms the former, then we shall have evidence that alleged errors of recall were not responsible for the effects.

Additional analyses, based on a sample of approximately 2,400 men enrolled in 140 colleges and universities, are shown in Table 1. These students were asked their LOA shortly after they entered college in Fall 1959, and again after they had completed 2 years of col-

[2] I called this ambiguity to Astin's attention in correspondence prior to the revision of his paper. Since he has not commented on the asynchronism in the objects of the "actual" and retrospective reports I assume this ambiguity cannot be ruled out.

TABLE 1

COVARIANCE ANALYSIS FOR COMPARING CHANGES IN ASPIRATION AMONG MEN EXPOSED TO DIFFERENT COLLEGE PRESS

Basis for classifying students into press groups [a]	Press reported by group with higher adjusted final LOA	*F* ratio
Faculty press		
Enthusiasm	strong	46.45**
Humanism	strong	25.42**
Affiliation (+) [b]	strong	23.58**
Independence (+)	strong	23.32**
Compliance	weak	19.88**
Vocationalism (−)	weak	8.48**
Achievement	strong	5.90*
Supportiveness (+)	strong	5.74*
Directiveness		.01
Student press		
Estheticism	strong	24.48**
Reflectiveness	strong	7.32**
Social conformity		.41
Competition		.23

[a] Results for comparisons based upon classification by faculty press for compliance and by student press for estheticism and reflectiveness should be interpreted with caution because of heterogeneity of regression among treatment groups.
[b] Direction of statistically significant relationships between press and increase in LOA found in earlier study.
* $p \leq .05$.
** $p \leq .01$.

lege, in June 1961.[3] Since these students were *not* asked to recall their earlier degree aspirations, but were asked about their current plans on two independent occasions, we can rule out errors of recall in the data. Revised forms of 9 of the 10 faculty press scales and 4 of the 10 student press scales used in the earlier study were included in the scales administered to students in the new sample. Only results for the 13 overlapping press scales are shown in Table 1.

It can be seen that the results confirm and extend the findings in my earlier report. First, each of the four kinds of faculty press (affiliation, independence, vocationalism, and supportiveness) previously associated with increase in LOA was also found to be significantly related in this analysis. The directions of the relationships are identical in both studies. On one scale (supportiveness) the significance level is lower, but on two (affiliation and vocationalism) the significance level is higher, compared to the earlier results. Second, both analyses agree in failing to reject the null hypothesis for three scales (directiveness, social conformity, and competition). Third, at least three additional significant relationships appeared; increase in LOA was associated with strong faculty press for enthusiasm, humanism, and achievement. Differences appearing on the remaining scales are ambiguous because of failure to meet the assumption of homogeneity of regression required for analysis of covariance.

Clearly the previously reported associations with faculty press did not disappear when errors of recall were ruled out; on the contrary, they appeared again, and were accompanied by additional faculty press associations. The fact that additional effects were suggested in this analysis could be due to a number of factors: increased reliability of the measures of intial LOA (as previously noted, an increase in the reliability of measurement should make it easier to discover effects), increased sample size, or the somewhat greater range of talent represented in this sample. In any case, it seems clear that these results clearly rule out the first of the rival hypotheses proposed by Astin.

APPROPRIATENESS OF COVARIANCE ANALYSIS

Astin raises two questions concerning the appropriateness of the method of

[3] This sample was drawn from a larger sample of students participating in the 1959 National Merit Scholarship Qualifying Test. The characteristics of the larger sample have been described elsewhere by Thistlethwaite (1960).

analysis in my earlier report: First, were the assumptions of covariance analysis satisfied? Second, was it proper to map the set of aspiration levels into an interval scale?

Three assumptions are commonly cited as prerequisities for use of simple linear covariance adjustment: linearity of regression, homogeneity of regression, and homogeneity of variance in the residual errors of estimate. Tests were made of the tenability of the last two of these assumptions in the analysis reported earlier. Except where noted these assumptions were satisfied. Since the consequences of departure from linearity of regression should be to overestimate the residual errors of estimate, and probably to underestimate the *F* ratios, it was simply assumed that the relationship between pretest and posttest LOA was linear. The heteroscedasticity noted by Astin probably means that the residual errors of estimate are not normally distributed. Cochran (1947) and Lindquist (1953) have discussed the effects of nonnormality upon the *F* tests used in analysis of variance, and the related work presumably carries over—for instance, similar conclusions probably apply for the effects of nonnormality in residual errors of estimate in covariance analysis (Cochran, 1957). The general import of this discussion is that the lack of normality does not preclude use of the *F* distribution, particularly if the tabled probabilities are doubled. Accordingly two-tailed tests of significance were used throughout in the earlier study, though most of the hypotheses were directional in nature.

A simple, dogmatic answer cannot be given to the second of the above questions. Some investigators prefer to map their data into an interval number system; others do not. Some justify mapping into a strong system of analysis on the basis of pragmatic results; others maintain that we should let the data determine the level of measurement employed. However, even if we grant that there are reasonable doubts about the appropriateness of mapping the LOA levels into an interval scale, Astin has not shown how such mapping could produce spurious results. A simple way of determining whether the results are specific to the method of analysis adopted is to repeat the comparisons using a different method of analysis.

RESULTS OBTAINED WITH AN ALTERNATIVE MODE OF ANALYSIS

Table 2 shows that the overall effects are very similar when we analyze the data by the alternative method used by Astin. In Table 2 mean press scores on each scale have been compared for subgroups which had the same pretest LOA but which differed in degree of change in LOA over a 2 year interval. To make these tests comparable to the overall *F* tests obtained in Table 1 (and in the earlier report) combined tests of significance were calculated by a method described by Mosteller and Bush (1954, p. 330) which permit one to weight the normal deviates in proportion to the sizes of the subgroups.

Both Tables 1 and 2 show that increase in LOA was associated with: (*a*) strong faculty press for enthusiasm, humanism, affiliation, independence, achievement, and supportiveness; (*b*) weak faculty press for compliance; and (*c*) strong student press for estheticism. The second analysis did not confirm the association with faculty press for vocationalism or with student press for reflectiveness, but did suggest a significant association with faculty press for directiveness.[4]

[4] Actually, the results on the scales for faculty press for vocationalism and student press for reflectiveness are confirmed if we use one-tailed tests of significance.

TABLE 2

COMPARISON OF AVERAGE PRESS SCORES OF MEN SHOWING DIFFERENT CHANGES IN ASPIRATION

Press scale	Sign and value of t for differences between means (mean for subjects having same or higher LOA in 1961 *minus* mean for subjects having lower LOA in 1961)[a]				
	Initial LOA in 1959				
	Bachelor's (1)	Master's (2)	Professional degree (MD, DDS, etc.) (3)	Doctorate (PhD or the equivalent) (4)	Combined test of significance z (5)
Faculty press					
Enthusiasm (+)[b]	+2.32*	+3.75**	+1.67	− .17	+4.00**
Humanism (+)	+4.32**	+4.17**	+4.19**	− .48	+6.50**
Affiliation (+)	+3.17**	+3.32**	+ .30	+ .11	+3.95**
Independence (+)	+2.88**	+5.16**	+2.57*	+ .05	+5.60**
Compliance (−)	−6.35**	−1.63	−1.97*	− .83	−6.40**
Vocationalism (−)	− .95	+ .05	−2.45*	− .65	−1.89
Achievement (+)	+1.98*	+2.47*	+2.02*	− .74	+3.13**
Supportiveness (+)	+2.36*	+2.30*	+2.14*	− .69	+3.39**
Directiveness (*ns*)	+ .52	+4.57**	+1.68	−2.73**	+2.26*
Student Press					
Estheticism (+)	+ .24	+3.58**	+2.54*	− .50	+2.67**
Reflectiveness (+)	+ .80	+1.41	+1.88	− .63	+1.78
Social conformity (*ns*)	+ .01	+1.21	+ .37	−1.02	+ .38
Competition (*ns*)	+3.33**	+ .73	+1.25	−4.31**	+1.85

[a] Signs in Column 1 are for the difference obtained by subtracting the mean for subjects having the same or lower LOA from the mean for subjects having higher LOA in 1961. *N*s for the groups having higher LOA in Columns 1–4 were 341, 398, 375, and 235, respectively; *N*s for groups having lower LOA in Columns 1–4 were 553, 161, 141, and 168, respectively.

[b] Direction of relationship between press and increase in LOA found in Table 1.

* $p \leq .05$.
** $p < .01$.

The individual t values show that faculty and student press were virtually unrelated to the changes in LOA of students who entered college with plans to seek the PhD. On the other hand, faculty and student press were highly related to changes in aspirations of students who entered college with plans to seek the MD or lower-level degrees.

It is clear that the effects reported in my earlier study were not specific to the method of analysis used; it may therefore be concluded that there is no evidence that those results were spurious due to my use of covariance procedures.

SOME SAMPLING CONSIDERATIONS

Finally, brief comment should be made on the statistical tests reported in Table 3 of Astin's paper. His subgroups have *N*s of 6 and 14, at the bachelor's level, and *N*s of 12 and 11 at the master's level. Now we have just seen the effects of faculty press appear primarily among students who do not initially plan to seek the PhD or its equivalent. It is hardly surprising that the null hypothesis was not rejected with such small samples particularly at the lower LOA levels. My earlier paper specifically called attention to the attenuation in the range of variation of the in-

dependent and dependent variables; it should also have been apparent from the epsilon values reported that I was dealing with relatively weak effects. Furthermore, Astin made no attempt to obtain combined tests of significance (over levels), so that the results would be comparable with the *F* ratios obtained in the covariance analysis. Consequently it is difficult to see how Astin could have expected the null hypothesis to be rejected in these comparisons.[6] In any case, his results are clearly less definitive than those which we have just considered in Table 2.

DISCUSSION

Since all of the studies so far reported on college productivity, or on changes in motivation for advanced training, have been based upon correlational data, there remains a reasonable doubt about the validity of *any* causal hypothesis advanced as an explanation for the observed relationships. I am disposed to explain these relationships as being due to the kinds of demands, images, and pressures which teachers and peers create, while Astin is disposed to see the explanation in terms of precollege characteristics. We can settle controversies about specific rival hypotheses (as I believe we have in this exchange), but there is of course no way of ruling out all possible conjectures about rival interpretations. Since the number of "input" variables is potentially infinite one can always interpret effects in terms of new, uncontrolled input characteristics. In the hope of obtaining some leads about methods of making teaching more effective I am willing to risk the temporary acceptance of one or more false hypotheses. However, nothing that Astin has presented convinces me he has yet identified any viable hypotheses that are genuine rivals to the hypothesis that learning environments shape in some degree the student's motivation for advanced training.

[6] A similar overreliance on small sample statistics is revealed in Astin's (1961) earlier paper, in which he has argued that precollege characteristics accounted for some of the relationships between faculty press and college productivity. The analysis reported there entailed second-order partial correlations based upon *N*s of 36. However, partial correlations are notoriously affected by errors in measurement and little confidence can be placed in the partials when the basic *r*'s are so unreliable.

REFERENCES

ASTIN, A. W. A re-examination of college productivity. *J. educ. Psychol.*, 1961, 52, 173–178.

ASTIN, A. W. Influences on the student's motivation to seek advanced training: Another look. *J. educ. Psychol.*, 1962, 53, 303–309.

COCHRAN, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, 3, 22–38.

COCHRAN, W. G. Analysis of covariance: Its nature and uses. *Biometrics*, 1957, 13, 261–281.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. New York: Houghton Mifflin, 1953.

MOSTELLER, F., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. Reading, Mass.: Addison-Wesley, 1954. Pp. 289–334.

THISTLETHWAITE, D. L. Recruitment of high ability youth in higher educational programs. Paper read at American Association for the Advancement of Science, New York, December 1960.

THISTLETHWAITE, D. L. Fields of study and development of motivation to seek advanced training. *J. educ. Psychol.*, 1962, 53, 53–64.

(Received July 9, 1962)

Journal of Educational Psychology
1962, Vol. 53, No. 6, 316–322

SEX, SOCIAL CLASS, AND ANXIETY AS SOURCES OF VARIATION IN SCHOOL ACHIEVEMENT[1]

BEEMAN N. PHILLIPS
University of Texas

The hypothesis that social class interacts with sex and anxiety to produce differences in school achievement was tested utilizing a sample of 759 adolescents classified into 8 subsamples involving 2 levels of anxiety and social class, and both sexes. 8 measures of school achievement were employed, including standardized tests and teacher grades. The results supported 2 major findings of previous research: females had higher anxiety scores than males, and highly anxious Ss had lower achievement and intelligence scores. Both 1st and 2nd order interactions between sex, social class, and anxiety were found. Difficulties were encountered in integrating these findings into anxiety theory, and certain proposals are offered to accomplish this.

Despite the great number of studies which have been reported, it is still difficult to reach definitive conclusions regarding anxiety and its relationships with school attitudes, behavior, and achievement. One of the reasons for this is that sex differences in the consequences of anxiety frequently have been reported (e.g., McGuire, Hindsman, King, & Jennings, 1961; Phillips, Hindsman, & McGuire, 1960; Sarason Davidson, Lighthall, Waite, & Ruebush, 1960). In addition, it has been found that sex interacts with a variety of factors in these differences in the consequences of anxiety (e.g., Grooms & Endler, 1960; Phillips, Hindsman, & Jennings, 1960; Reese, 1961). Finally, no theory of anxiety seems capable of explaining these multiple results.

In this situation it appears that a further step which needs to be taken is the investigation of factors which have relationships with sex, school achievement, and anxiety. One such factor is social class, which is clearly related to all three; and it was the purpose of this study to investigate social class, sex, and anxiety as interrelated factors in school achievement. The investigation was designed to test a general hypothesis, which was that differences in school achievement would be associated with interactions of sex, social class, and anxiety. In addition, the implications of such findings with regard to prevailing conceptualizations of anxiety, and its relationships with school achievement, were analyzed.

METHOD

The 759 subjects of this study were selected from approximately 1,500 seventh grade students in four Texas communities who form the total population of the University of Texas Human Talent Project. The instruments used included the California Test of Mental Maturity (CTMM); the Language Arts, Mathematics, and Reading subtests of the California Achievement Test (CAT); the Social Studies and Science subtests of the Sequential Test of Educational Progress (STEP); and the Castaneda-McCandless Anxiety Scale, with items adapted for use with adolescents. An index of social class based on source of income, occupation, and education of the parents also was computed for each subject (McGuire et al.,

[1] The research reported here was performed pursuant to OE-2-10-003 with the United States Office of Education, Department of Health, Education, and Welfare. Also, the author wishes to express his appreciation to G. McBee who supervised IBM 650 analyses of data.

1961). In addition, teacher grades were obtained for the subject areas of language arts, mathematics, and social studies at the end of the school year. All of these variables were linearly transformed to stanine values so that each had a mean of 5 and a standard deviation of 1.5.

Basically a fixed $2 \times 2 \times 2$ design was utilized with two levels of social class and anxiety, and two sexes. In order to achieve a clearer differentiation between subsamples, the subjects with stanines of 5 on social class or anxiety were not included in the analyses. Therefore, the "high" levels of social class (middle class) and anxiety included subjects with stanine values of 6 or greater, and the "low" levels of social class (lower class) and anxiety included subjects with stanine values of 4 or less. Thus, there were a total of 8 subsamples drawn from the total population. Another factor, type of achievement measure with two categories, teacher grades and objective tests, was added to the basic design.

Two different approaches to school achievement were employed. The first was based on the *relationship* between intelligence and achievement, which was determined by computing zero-order correlation coefficients between total IQ and each of the eight achievement scores, and then performing an analysis of variance of Fisher z transformations of these correlations.

The other approach was based on the *level* of achievement reached, and for this means and standard deviations were computed on all achievement variables, and t tests and an analysis of variance were performed on these means.

In addition, chi square analyses were made of the distributions of high and low anxiety scores; intelligence means were computed and tested for significance; and differences between intelligence and achievement means were determined.

RESULTS

The analyses of the data are presented in the following order. The results with respect to the correlations are given in Tables 1 and 2. Table 1 gives the actual coefficients and Table 2 shows the result of the analysis of variance of Fisher z transformations. Then in Table 3, the results with respect to the mean differences in achievement between the subsamples are given, while Table 4 presents an analysis of variance of these means. Finally, Table 5 gives the IQ means and standard deviations of the subsamples; Table 6 presents a chi square analysis of the sex and social class distribution of the high and low anxiety scores; and Table 7 shows differences between mean intelligence and mean achievement scores.

TABLE 1

CORRELATIONS OF IQ WITH ACHIEVEMENT VARIABLES

	Males				Females			
	High social class		Low social class		High social class		Low social class	
Achievement variables	High anxiety	Low anxiety	High anxiety	Low anxiety	High anxiety	Low anxiety	High anxiety	Low anxiety
Teacher grades								
Social studies	54	57	29	29	42	37	35	33
Mathematics	66	60	36	21	29	32	36	36
Language arts	54	58	30	24	40	38	34	30
Objective tests								
STEP Social Studies	56	63	60	47	49	57	53	56
CAT Mathematics	69	61	50	34	45	43	43	51
CAT Language Arts	70	63	59	53	41	51	49	52
CAT Reading	70	63	63	51	55	66	52	63
STEP Science	50	57	51	38	49	53	44	47
N	55	99	108	103	58	94	158	84

Note.—Decimal points have been omitted.

TABLE 2

ANALYSIS OF VARIANCE OF FISHER z TRANSFORMATIONS OF CORRELATION COEFFICIENTS

Source	*SS*	*df*	*F*
Sex	.17	1	75.09**
Social class	.22	1	96.17**
Anxiety	.00	1	—
Achievement	.56	1	244.39**
Sex × social class	.19	1	81.65**
Sex × anxiety	.04	1	17.09**
Sex × achievement	.00	1	—
Social class × anxiety	.01	1	2.30
Social class × achievement	.07	1	29.04**
Anxiety × achievement	.00	1	—
Sex × social class × anxiety	.00	1	—
Sex × social class × achievement	.04	1	18.61**
Sex × anxiety × achievement	.02	1	7.87**
Social class × anxiety × achievement	.00	1	—
Sex × social class × anxiety × achievement	.00	1	—
Within	.11	48	—

** $p < .01$.

Now, summarizing the results of Table 2, it may be stated that:

1. Three main effects were statistically significant: IQ was more highly correlated with achievement for males than for females, IQ was more highly correlated with achievement in the middle class than in the lower class, and IQ was more highly correlated with objective tests than with teacher grades.

2. Three first order interactions also were statistically significant: an increase in social class resulted in a higher correlation of IQ with achievement for males than for females, an increase in anxiety resulted in an increase in the correlation of intelligence with achieve-

TABLE 3

SUBSAMPLE MEANS AND STANDARD DEVIATIONS FOR THE ACHIEVEMENT VARIABLES

	Males								Females							
	High social class				Low social class				High social class				Low social class			
	High anxiety		Low anxiety		High anxiety		Low anxiety		High anxiety		Low anxiety		High anxiety		Low anxiety	
Achievement variables	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*	$\bar{X}$	*SD*
Teacher grades																
Social studies	4.9	1.4	5.3	1.3	4.7*	1.4	3.9	1.7	5.2*	1.3	6.3	1.1	4.0*	1.5	5.3	1.4
Mathematics	4.8	1.5	5.2	1.3	4.7*	1.6	4.0	1.6	5.2*	1.3	6.1	1.3	4.2*	1.5	5.2	1.5
Language arts	4.7	1.4	5.1	1.2	4.6*	1.3	3.8	1.5	5.3*	1.0	6.3	1.0	4.5*	1.5	5.4	1.3
Objective tests																
STEP Social Studies	5.6	1.6	5.5	1.6	4.6	1.3	4.3	1.0	5.6*	1.1	6.4	1.4	4.5*	1.2	4.9	1.2
CAT Mathematics	5.3	1.5	5.2	1.5	4.5*	1.3	4.1	1.2	5.9*	1.1	6.4	1.4	4.6*	1.5	5.2	1.4
CAT Language	5.1	1.5	5.2	1.5	4.5	1.5	4.2	1.4	5.3*	1.2	5.9	1.4	4.2*	1.5	5.0	1.4
CAT Reading	5.5	1.6	5.6	1.5	4.6	1.4	4.3	1.3	5.8*	1.1	6.3	1.3	4.4*	1.3	4.9	1.3
STEP Science	5.6	1.6	5.6	1.7	4.7	1.5	4.4	1.3	5.4*	1.4	6.2	1.1	4.4*	1.3	4.8	1.3

* Difference between this high anxiety subsample mean and the low anxiety subsample mean (in the next column) statistically significant at or beyond the .05 level.

TABLE 4

ANALYSIS OF VARIANCE OF ACHIEVEMENT VARIABLE MEANS

Source	*df*	*MS*	*F*
Sex	1	23947.0	45.9**
Social class	1	154842.0	296.6**
Anxiety	1	5112.0	9.8**
Sex × social class	1	1090.0	2.1
Sex × anxiety	1	27890.0	53.4**
Social class × anxiety	1	3813.0	7.3**
Sex × social class × anxiety	1	8882.0	17.0**
Within[a]	56	522.0	

[a] Based on the eight subsamples and the eight mean achievement scores.
** $p < .01$.

TABLE 5

SUBSAMPLE IQ MEANS AND STANDARD DEVIATIONS

Subsample	$\overline{X}$	*SD*
Male		
High social class, high anxiety	5.7	1.5
High social class, low anxiety	5.8	1.5
Low social class, high anxiety	4.8*	1.4
Low social class, low anxiety	4.3	1.2
Female		
High social class, high anxiety	5.5*	1.4
High social class, low anxiety	6.2	1.3
Low social class, high anxiety	4.3*	1.3
Low social class, low anxiety	4.8	1.3

* Difference between this mean and the mean directly below it statistically significant at or beyond the .05 level.

TABLE 6

CHI SQUARE ANALYSES OF DISTRIBUTION OF ANXIETY SCORES BY SEX AND SOCIAL CLASS

Subsample	Anxiety: High	Anxiety: Low	Chi square
High social class			
Males	55	99	—
Females	58	94	
Low social class			
Males	108	103	9.40**
Females	158	84	
Males			
High social class	55	99	8.64**
Low social class	108	103	
Females			
High social class	58	94	27.88**
Low social class	158	84	

** $p < .01$; $df = 1$.

ment for males but a decrease in this correlation for females, and an increase in social class resulted in a much higher increase in the correlation of IQ with objective tests.

3. Finally, two second order interactions were statistically significant: for males, an increase in social class resulted in a greater increase in the correlation of IQ and objective tests, while for females, an increase in social class resulted in an increase in the correlation of IQ and teacher grades, but there was no change in the correlation of IQ and objective tests; and for males, an increase in anxiety resulted in an increase in the correlations of IQ and teacher grades, and IQ and objective tests, while for females, an increase in anxiety resulted in an increase in the

TABLE 7

SUBSAMPLE DIFFERENCES BETWEEN MEAN IQ AND THE MEAN OF THE ACHIEVEMENT MEANS

Subsample	Teacher grades[a]	Objective test[b]
Male		
High social class, high anxiety	−.9*	−.3
High social class, low anxiety	−.6*	−.4
Low social class high anxiety	−.1	−.2
Low social class low anxiety	−.4*	0
Female		
High social class, high anxiety	−.3	.1
High social class, low anxiety	0	0
Low social class, high anxiety	−.1	.1
Low social class, low anxiety	−.5*	.2

[a] Difference obtained by determining the mean of the three teacher grade means and subtracting the mean IQ.
[b] Difference obtained by determining the mean of the five objective test means and substracting the mean IQ.
* Difference is statistically significant at or beyond the .05 level.

correlation of IQ and teacher grades but a decrease in the correlation of IQ and objective tests.

Referring, now, to Table 4, the following results may be noted:

1. Three main effects were statistically significant: females had higher achievement than males; middle-class subjects had higher achievement than lower-class subjects; and subjects with low anxiety had higher achievement than subjects with high anxiety.

2. Three interactions were statistically significant: an increase in anxiety resulted in lower achievement for females and slightly higher achievement for males; an increase in anxiety resulted in lower achievement for middle-class subjects but had no appreciable effect on the achievement of lower-class subjects; and an increase in anxiety resulted in an increase in the achievement of lower-class males and a decrease in the achievement of lower-class females, while an increase in anxiety resulted in a larger decrease in the achievement of middle-class females than it did in middle-class males.

Next, in reference to Tables 5 and 6, two major findings may be reported: first, an increase in anxiety among females resulted in lower intelligence scores in both social classes, while an increase in anxiety among males resulted in higher intelligence scores in the lower class, and made little difference in the middle class; second, there were proportionately more males and females with high anxiety in the lower class than in the middle class, although in the lower class a larger proportion of females than males had high anxiety.

Finally, from the results of Table 7, it was found that: for males in the middle class, mean teacher grades were lower than mean intelligence in both high and low anxiety groups; for males in the lower class, mean teacher grades were lower than mean intelligence in only the low anxiety group; for females in the middle class, there were no differences between mean teacher grades and mean intelligence in either high or low anxiety groups; for females in the lower class, mean teacher grades were higher than mean intelligence in the low anxiety group, with no mean difference in the high anxiety group; and lastly, differences between standardized achievement and intelligence were insignificant in all subsamples.

DISCUSSION

Viewed in relation to the overall findings of past anxiety research, it is evident that the general results of this study were in agreement on two major points. Sex differences in level of anxiety were obtained, with males having lower anxiety scores than females; and differences in the achievement of subjects with high and low anxiety were found, with high anxiety being associated with lower achievement. However, the significance of these areas of agreement is diminished by the finding that both the level of anxiety and its relationship to school achievement depended on the interaction of sex and social class.

As the first indication of this interaction, sex differences in level of anxiety occurred *only* in the lower social class, where females had significantly higher anxiety scores than males. In addition to this, it should be noted that middle-class and lower-class males differed less in anxiety scores than middle-class and lower-class females.

As a further indication of the interaction of sex and social class, lower-class males with high anxiety had higher intelligence scores than lower-class males with low anxiety, although middle-class males with high anxiety did not differ

significantly in intelligence from middle-class males with low anxiety. Furthermore, in both social class levels females with high anxiety had lower intelligence scores than females with low anxiety.

Finally, the interaction between sex and social class was evident in relationships between anxiety and school achievement. In general, anxiety as a debilitating factor was more evident in relation to teacher grades than in relation to achievement tests, with the achievement of lower-class males being less affected than the achievement of other adolescents.

Considering what these findings may mean in terms of anxiety theory, Sarason, Davidson, Lighthall, and Waite (1958) have speculated that males obtain lower anxiety scores than females because they are more reluctant to admit anxiety. As they stated it, anxiety is "ego-alien" to males and "ego-syntonic" to females. In other words, the expression of anxiety is incongruous to the male's conception of himself, and this causes him to be more defensive about admitting anxiety. Accepting this, the results of this study may be explained by assuming that social class is a factor in whether anxiety becomes ego-alien or ego-syntonic. More specifically, the greater differentiation of sex roles in the lower class, particularly with respect to masculinity, causes the ego-alien aspects of anxiety to be more highly developed and delineated in lower-class males.

Although sex and social-class differences in level of anxiety seem to fit into the ego-alien and ego-syntonic rationale, more difficulty is experienced in trying to explain the role of sex and social class in relationships between anxiety and school achievement. According to the rationale, anxiety should have greater disruptive effects on males, due to the greater ego-alien nature of anxiety in males. But, as previously noted, anxiety appeared to interfere less with achievement of males than females, especially in the lower class. Thus, whether anxiety produces responses which are compatible or incompatible with school achievement tasks seems to depend upon social class. More to the point, the social class related experiences of lower-class males appear to be especially significant in whether anxiety has facilitative effects.

Finally, investigators espousing drive conceptions of anxiety have been particularly interested in the performance of highly anxious and less anxious subjects on complex and simple tasks. According to the Hull-Spence drive theory, anxiety should interfere with performance on complex tasks, and if tasks on intelligence tests can be considered complex, the negative relationships usually obtained between anxiety and intelligence appear to be consistent with this view. However, the results of this study did not altogether conform to the prediction of drive theory since lower-class males with high anxiety had *higher* intelligence scores than lower-class males with low anxiety. Apparently, then, the relationship between anxiety and intelligence is not uniformly negative across samples differentiated in terms of sex and social class, which raises the question of whether these factors in some way influence the drive properties of anxiety.

In summary, it is apparent that sex differences are not easily integrated into prevailing theories about anxiety, and that the interaction of sex, social class, and anxiety represents a particular challenge to existing theoretical formulations. Furthermore, it is argued that an integration of research findings concerning anxiety and ego involvement, and social class and school achievement, offers the best opportunity for a con-

ceptualization which will successfully handle such differences. Toward this end, the following facts and speculations are presented:

1. It is known that middle-class subjects have a greater desire to do well in school than lower-class subjects, and that a similar difference exists between females and males, with this difference being greater in the lower class. As a result, school achievement tasks should be more ego involving to middle-class subjects than to lower-class subjects, and to females more than males, especially in the lower class.

2. There have been studies of the relationship between anxiety and performance under neutral and ego involving conditions, and the gist of these studies is that highly anxious subjects perform better under neutral conditions, while less anxious subjects perform better under ego involving conditions. Therefore, the degree to which anxiety interferes with a subject's performance in a situation depends upon the degree to which the subject is ego involved.

3. Integrating these findings, since school achievement tasks are more ego involving to middle-class subjects, and to females more than males (particularly in the lower class), the degree to which anxiety interferes with school achievement should depend upon both the sex and the social class status of subjects. Taking into account these specific relationships, anxiety should interfere the least with the school achievement of lower-class males, and this is what the overall findings of this study showed.

REFERENCES

GROOMS, R. R., & ENDLER, N. S. The effect of anxiety on academic achievement. *J. educ. Psychol.*, 1960, **51**, 299–304.

MCGUIRE, C., HINDSMAN, E., KING, F. J., & JENNINGS, E. Dimensions of talented behavior. *Educ. psychol. Measmt.*, 1961, **21**, 3–38.

PHILLIPS, B., HINDSMAN, E., & JENNINGS, E. Influence of intelligence on anxiety and perception of self and others. *Child Develpm.*, 1960, **31**, 41–46.

PHILLIPS, B., HINDSMAN, E., & MCGUIRE, C. Factors associated with anxiety and their relation to the school achievement of adolescents. *Psychol. Rep.*, 1960, **7**, 365–372.

REESE, H. W. Manifest anxiety and achievement test performance. *J. educ. Psychol.*, 1961, **52**, 132–35.

SARASON, S. B., DAVIDSON, K., LIGHTHALL, F., & WAITE, R. Classroom observation of high and low anxious children. *Child Develpm.*, 1958, **29**, 287–296.

SARASON, S. B., DAVIDSON, K. S., LIGHTHALL, F. K., WAITE, R. R. & RUEBUSH, B. K. *Anxiety in elementary school children.* New York: Wiley, 1960.

(Received June 21, 1962)

4 MODERN EDUCATIONAL PSYCHOLOGY TEXTS FROM MACMILLAN

EDUCATIONAL PSYCHOLOGY, Second Edition

By Glenn M. Blair, R. Stewart Jones, Ray H. Simpson, all of the University of Illinois

This text retains the functional orientation of the first edition, carefully illustrating psychological theories with actual classroom examples and case studies. Emphasis is on the total development and adjustment of the child, clearly outlining both his needs and the forces that motivate his learning. Additional material has been included on teaching machines, feedback, "brainstorming," creative thinking, and discovery and learning.

678 pages **$7.00**

READINGS IN EDUCATIONAL PSYCHOLOGY

Edited by Victor H. Noll and Rachel P. Noll both of Michigan State University

This book of readings, designed to accompany standard texts for introductory courses in educational psychology, seeks to broaden the interests and understanding of undergraduate students. Selections have been included on growth and development, learning, measurement and evaluation, mental hygiene, and the science of educational psychology. Introductory comments place each selection in its appropriate setting and highlight the contribution of the work under discussion.

608 pages **$3.95**

COMING IN THE SPRING—

PERSONALITY AND BEHAVIOR

By Jesse E. Gordon, University of Michigan

This textbook provides a detailed exploration and analysis of the formation and operation of behavioral patterns in personality. The first part of the book explains clearly the essential basis of human behavior, developing such topics as physiological functioning, the basic laws of learning, and motivation. The second section describes the ways in which various factors interact and combine to produce the great variety of individual personality patterns.

DYNAMICS OF ADOLESCENT ADJUSTMENT

By Thomas F. Staton, Huntingdon College, Montgomery, Alabama

This book is designed to help students better understand adolescents, what is happening to them, how they think and feel, why they act as they do, how to recognize when they need help, and how to help them. The decision to include a particular theory, concept, or interpretation has been based on the extent to which that factor has proved true and useful in the author's own clinical experience.

THE MACMILLAN COMPANY 60 Fifth Avenue, New York 11, N.Y.

Publication Manual

of the
American
Psychological
Association

1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference list are included. Scientists writing for publication will find the Publication Manual an indispensable guide.

Price, $1.00

Order from

**AMERICAN
PSYCHOLOGICAL
ASSOCIATION**

**Order Department
1333 Sixteenth Street, N. W.
Washington 6, D. C.**

ELEMENTARY STATISTICAL METHODS IN PSYCHOLOGY AND EDUCATION

Paul Blommers and E. F. Lindquist

This book encourages the beginning college student in the uses and interpretation of statistics, stressing the importance of a critical evaluative attitude. The text explores in depth a limited number of basic statistical concepts and techniques, with emphasis on the logico-mathematical basis of statistics for the student with limited mathematical background. A Study Manual accompanies the text, and an Instructors' Key is available.

528 pages *1960* *$6.50*

PSYCHOLOGY IN TEACHING AND LEARNING

William Clark Trow

This compact text relates the major theories of educational psychology to the daily activities of the teacher. Each chapter introduces an educational situation and the psychological theories and facts needed to understand and handle it. Student's and Instructor's Manuals.

488 pages *1960* *$5.75*

Houghton Mifflin

BOSTON · NEW YORK · ATLANTA
GENEVA, ILL. · DALLAS · PALO ALTO

Chandler Publications in Educational Psychology
DAVID G. RYANS, Editor

Statistical Measures:
A PROGRAMMED TEXT

by FRANK F. GOROW, Long Beach State College

THIS scrambled textbook contains branching programs designed to enable the student to learn to compute simple statistical measures on his own, outside regular class hours. It is not proposed as a text in statistics, but as an adjunct for courses such as tests and measurements in which ability to compute simple statistical measures is an essential skill, but is not central to the subject matter of the course. The programs have been thoroughly tested through several preliminary editions by the author and others. Gorow reports that he now uses one class period to cover this material in comparison to six devoted to these topics before development of the programs.

A manual containing the program diagrams is available to teachers.

192 pp., paper, $2.25

ABBREVIATED CONTENTS

Part One: SMALL SETS OF SCORES—The Mean, The Median, The Mode, Three Measures of Central Tendency, the Range, the Standard Deviation, the Relation of the Measures, General Assignment.

Part Two: LARGE SETS OF SCORES—Grouping Data in a Frequency Distribution, Computing the Median from a Frequency Distribution, Computing Quartiles, Deciles, and Percentiles, Simple Measures of Dispersion, Computing the Mean, Computing the Standard Deviation, General Assignment.

Part Three: MEASURES OF RELATIONSHIP—The Concept of Correlation, The Correlation Scattergram, The Coefficient of Correlation, General Assignment.

Appendices: Interpretation of Test Data, A Few Useful References, Summary and Formulas.

Also in this series:

BASIC CONCEPTS OF TEACHING: With Brief Readings, *cloth,* $4.25
BASIC CONCEPTS OF TEACHING: Concise Edition, *paper,* 2.25

both by ASAHEL D. WOODRUFF, *University of Utah*

Teachers wishing to consider these books for adoption are invited to write for examination copies.

CHANDLER PUBLISHING COMPANY
604 Mission Street, San Francisco 5